اُردو اطلاعیات

(جلد اوّل)

ڈاکٹر عطش دُرّانی

مقتدرہ قومی زبان
پاکستان

اُردو اطلاعیات

(جلد اوّل)

ڈاکٹر عطش دُرّانی

مرکزِ فضیلت برائے اُردو اطلاعیات

مقتدرہ قومی زبان۔ پاکستان

۲۰۰۸ء

سلسلہ مطبوعات مقتدرہ: ۴۸۵

عالمی معیاری کتاب نمبر ISBN ۹۷۸-۹۶۹-۴۷۴-۲۱۸-۲

طبع اوّل ۲۰۰۸ء

تعداد ۵۰۰

قیمت .... ۲۲۰/= روپے

فنی تدوین .... تنویر فاطمہ

(معاون افسر اطلاعیات، مرکز فضیلت)

طابع .... ورڈ میٹ پرنٹرز، اسلام آباد

ناشر .... **محمد اکرام بلال**

معتمد

مقتدرہ قومی زبان، ایوانِ اُردو،

پطرس بخاری روڈ، ایچ۔۸/۴،

اسلام آباد، پاکستان۔

فون: ۰۵۱-۹۲۵۰۳۱۱-۱۳

فیکس: ۰۵۱-۹۲۵۰۳۱۰

☆

عرضِ ناشر

یہ کریڈٹ مقتدرہ قومی زبان کا ہے کہ اُردو کو کمپیوٹر کی بنیادی زبان بنانے اور معیار بندی کرنے، سافٹ ویئر اُردو میں تبدیل کرنے اور اُردو اطلاعیات کی بنیادیں استوار کرنے کے لیے پہلا قدم اٹھایا۔ ۱۹۹۸ء سے ۲۰۰۸ء تک ان کاوشوں کے دس برس ایک ترقیاتی عشرے کی حیثیت رکھتے ہیں۔ اب ٹیکنالوجی کے میدان میں ہر طرف اُردو کا شُہرہ ہے۔ بہت سا کام "اخبارِ اُردو" جیسے جریدے کے ذریعے بھی پیش ہوتا رہا۔ یوں بہت سی خبریں، اطلاعات، مضامین اور تحقیقی نتائج اُردو اطلاعیات کی سمت بندی کرتے رہے ہیں۔

اُردو اطلاعیات کے حوالے سے یہ تمام عرصہ ڈاکٹر عطش دُرّانی کی بصیرت، مسلسل اور اَن تھک محنت، عزم، منصوبہ بندی اور رہنمائی کا ثمرہ ہے۔ اسی کے اثرات کئی اداروں، یونیورسٹیوں اور یونی کوڈ، آئی ایس او، مائیکروسافٹ اور موبائل فون کمپنیوں تک پڑے اور ہر طرف اُردو کے تکنیکی استعمال کا آغاز ہوا۔ یہ سب کچھ اسی سمت کی طرف اشارہ کناں ہے، جدھر مرکزِ فضیلت برائے اُردو اطلاعیات کی منصوبہ بندی کی گئی ہے اور جو بہت جلد اطلاعیات کے انسٹی ٹیوٹ کا روپ دھار لے گی۔

اس دوران میں بہت سے مقالات، مضامین، رپورٹیں اور پیش کاریاں سامنے آتی رہیں۔ یہ سب اُردو اطلاعیات کی اُس تاریخ کا ایک حصہ اور مستقبل کی بنیادیں ہیں، ڈاکٹر عطش دُرّانی نے جس کی نیو رکھی تھی۔ ضروری ہے کہ انھیں شائع کر کے محفوظ کر دیا جائے۔ چنانچہ پہلی جلد اُنھی کی نگارشات پر مبنی شائع کی جا رہی ہے۔ اگلی جلدیں دیگر ماہرین، اساتذہ، طلبہ اور ارکانِ عملہ کی تحریروں پر مشتمل ہیں۔ ان سب سے "اُردو اطلاعیات" کا مضمون (Discipline) تشکیل پاتا ہے۔ اس "کل" کے لیے یہ "آج" بہت ہی پائیدار اور مستحکم ہے۔ یہ جلدیں اسی توقع اور اُمید پر پیش کی جا رہی ہیں کہ اِنھی سے اُردو اطلاعیات کا مستقبل تشکیل پائے گا۔

(پروفیسر فتح محمد ملک)
صدر نشین

فہرست

پیش گفتار

"اُردو اطلاعیات ہی اُردو کا مستقبل ہے"۔ یہ جملہ ۱۹۹۸ء میں وجود میں آیا لیکن اس کی بنیادیں بہت پہلے رکھ دی گئی تھیں۔ ۱۹۸۰ء میں احمد مرزا جمیل نے تصویری کتابت "نوری نستعلیق" متعارف کرائی جو کمپیوٹر کے ذریعے عمل کار ہوئی۔ ۱۹۸۴ء میں مجھے مقتدرہ قومی زبان میں اس ڈیسک کے فرائض تفویض ہوئے کہ تکنیکی سطح پر اُردو کے استعمال کے امکانات پیشِ نظر رکھے جائیں۔ کمپیوٹر پر اُردو عمل کاری کی ایک آدھ نمایش بھی ہوئی۔ ایک تکنیکی کمیٹی کے اجلاس بھی منعقد ہوئے لیکن بات انفرادی اور ذاتی مفادات سے آگے نہ بڑھی۔

۱۹۹۸ء قومی زندگی کا وہ یادگار سال ہے، جب ایک طرف پاکستان نے ایٹمی دھما کا کیا اور دوسری طرف "اُردو اطلاعیات" کی نیو ڈالی گئی۔ نیشنل ڈیٹا بیس آرگنائزیشن، اسلام آباد (NDO) ان دنوں قومی شناختی کارڈ کی تیاری کے لیے پَر تول رہا تھا۔ لیکن اس وقت تک اُردو لفظ کاری (Word Processing) کے جتنے بھی سافٹ ویئر سامنے آئے تھے، وہ بنیادی طور پر شبیہ (Image) کے اُصول اور اپنے اپنے ایسکی ضابطے (ASCII Code) پر عمل کرتے تھے اور ان میں عمل کاری (Processing) صرف انھی کے سافٹ ویئر میں ہو سکتی تھی۔ کوئی بنیادی معیار بندی موجود نہ تھی۔ اسی طرح حرف کی بنیاد اور یونی کوڈ (UNICODE) یا آئی ایس او (ISO) جیسے معیارات کا دور دور تک پتا نہ تھا۔ انھی دنوں میں کچھ طالع آزما اٹھے تاکہ NDO کا کاروبار ان کے ہاتھوں میں آسکے۔ انھوں نے معیار بندی کی کوششوں کا آغاز کیا۔ یہی ادارہ NDO آگے چل کر NADRA کہلایا اور اُردو شناختی کارڈ ممکن ہوئے۔

معیار بندی کا پہلا ایسا سیمینار ستمبر ۱۹۹۸ء میں نیشنل یونیورسٹی فاسٹ (FAST) لاہور میں منعقد ہوا۔ مجھے اس میں شریک ہونے کا موقع ملا۔ صدارت پر بھی براجمان ہوا۔ مقامی کمیٹی کا انچارج بنا لیکن معیار بندی کے لیے کوئی راستہ نہ نکلا۔ اگلے اجلاس مقتدرہ قومی زبان میں شروع کیے۔ یہاں ستمبر ۱۹۹۹ء میں شعبہ "اُردو اطلاعیات" کی بنیاد رکھی گئی اور یوں معیاری کلیدی تختے اور معیاری ایسکی کوڈ پلیٹ کی تیاری کا کام

انجام پایا۔ اس وقت کی تمام کمیٹیوں کے ارکان رضا کارانہ کام کرتے تھے۔ اپنے اپنے اخراجات پر آتے جاتے رہتے۔ بالآخر ۵۔جون ۱۹۹۹ء کو ایسکی کوڈ پلیٹ کے ورژن 1.1 کا تعارفی سیمینار مقتدرہ قومی زبان کی طرف سے منعقد کیا۔ اس کی صدارت ڈاکٹر قدیر خان نے کی اور مہمان خصوصی وفاقی وزیر یاسین وٹو تھے۔ کابینہ سیکرٹری اے ڈبلیو قاضی آخری اجلاس میں موجود تھے۔ کلیدی تختہ حتمی صورت میں تعدّد شماری کی بنیاد پر سامنے آیا۔ نادرا میں آزمایا گیا اور بالآخر ۱۴۔دسمبر ۱۹۹۹ء کو جاری کر دیا گیا۔ اس کلیدی تختے پر دیگر پاکستانی زبانوں کے اداروں سے اتفاق رائے حاصل کیا گیا۔ اس مرحلے پر نادرا کے میجر سہیل اقبال، کابینہ سیکرٹری ڈاکٹر معصومہ حسن اور ایڈیشنل سیکرٹری جناب محمد اسماعیل پٹیل کا ذکر توجہ چاہتا ہے۔

ایسکی کوڈ پلیٹ کے اجراء میں چند اغلاط موجود تھیں نیز جب میں نے یہ دیکھا کہ بعض علامتوں اور پاکستانی زبانوں کی شرکت ایک ہی ضابطہ تختی اور ایک ہی فانٹ سے ممکن ہے تو اپنا نقطہء نظر "خالی کشتیوں" (Ghost Characters) کی صورت میں سامنے لایا۔ اس مقام پر جانب سید عمار جعفری کی خدمات اہم ہیں۔ مختلف ذیلی کمیٹیوں کی روداد یں اس کتاب کی دوسری جلد میں شامل ہیں۔

انھی دنوں کابینہ سیکرٹری ڈاکٹر معصومہ حسن نے اس کام کی قومی اہمیت دیکھ کر کابینہ کے اجلاس میں منظور کرا کے معیار بندی کا یہ کام سرکاری طور پر کابینہ ڈویژن کے ذمے لے لیا اور نئے کلیدی تختے اور نئے ضابطہ تختی ورژن ۲ کی بنیاد پر پہلا اُردو سافٹ ویئر مقابلہ منعقد کرانے کا حکم دیا۔ اس مقابلے میں حصہ لینے والوں کے کام ITCN Asia 2001 میں نمایش کے لیے پیش کیے۔ کابینہ سیکرٹری کی کوششوں سے ایسکی کوڈ پلیٹ (ضابطہ تختی) اور مقتدرہ کے لیے ڈبل ایسکی کوڈ کے عالمی معیار بندی کے ادارے یونی کوڈ کی رکنیت حاصل کرنے کی منظوری صدر پاکستان نے دے دی۔ اس موقع پر جناب طاہر مفتی اور جناب محمد اویس احمد کی خدمات قابل ذکر ہیں۔

اسی نمایش میں مائیکروسافٹ دوبئی آفس کے انچارج جناب احمد عبداللہ سے ملاقات ہوئی تو انھیں ایم ایس سافٹ ویئر اُردو میں بدلنے اور اُردو کلیدی تختہ استعمال کرنے کا مشورہ دیا۔ ہومیوپیتھی کا اُردو میں ہونا بطور مثال پیش کیا۔ مشورہ صائب تھا۔ مائیکروسافٹ نے Windows XP 2000 میں جزوی طور پر یہ کلیدی تختہ شامل کر لیا تھا اور ۲۰۰۲ء میں LLP کے تحت اپنے سافٹ ویئر اُردو میں بدلنے کا آغاز کیا اور یہ کام مقتدرہ ہی کے سپرد کیا۔ اس منصوبے کے تحت ہم نے Windows XP 2000 اور Office 2003 کو اُردو میں بدلا اور دعوی کیا کہ ہم نے کمپیوٹر سکرین اُردو میں بدل دی ہے۔ ان معاملات میں مائیکروسافٹ کے LLP کے انچارج جناب اینڈی عبرا اور پاکستان میں جناب وقار خیسانی کی خدمات بے مثل ہیں۔

اس دوران میں ایک PC-I حکومت کو پیش کیا گیا۔لیکن اس عرصے میں میری بجائے پاکستان کمپیوٹر بیورو کے سعید احمد کو مقتدرہ میں اُردو اطلاعیات کا انچارج بنا دیا گیا۔ یہ منصوبہ جہاں ایک طرف کابینہ ڈویژن سے پیش ہوا اور مسترد ہوا، وہیں دوسری طرف آئی ٹی ڈویژن کی طرف سے بھی چار کروڑ روپے کا یہی منصوبہ متوازی طور پر منظور کرالیا گیا اور ذیلی ٹھیکے میں اسے نیشنل یونیورسٹی فاسٹ (FAST) لاہور کو دے دیا گیا۔ وہاں کچھ کام ہوا یا نہیں لیکن ان کے اساتذہ اور طلبہ نے جو تحقیقی قدم اٹھائے، اُن سے اکثر "اخبارِ اُردو" کی زینت بنتے رہے۔

نئے کابینہ سیکرٹری جناب اعجاز رحیم نے جب مقتدرہ میں کمپیوٹر سکرین اُردو میں بدلتے اور میری ٹیم کو دن رات کام کرتے دیکھا تو اس کے نئے منصوبے "مرکزِ فضیلت برائے اُردو اطلاعیات" کو حکومت کی طرف سے منظوری دلا دی۔ اس کا پہلا مرحلہ کامیابی سے گزرا تو پاک نستعلیق فانٹ کی نمائش قومی اسمبلی اور سینیٹ کی قائمہ کمیٹیوں میں کی گئی۔ وزیراعظم پاکستان نے رپورٹ دیکھ کر کابینہ سیکرٹری کو یہ کام تیز تر کرنے کا حکم جاری کیا اور یوں یہ منصوبہ دوسرے مرحلے میں داخل ہوا۔ ۱۱ تا ۱۲۔مئی ۲۰۰۸ء کو یونی کوڈ کی ٹیکنیکل کمیٹی کے اجلاس میں Ghost Characters Theory کو کلی طور پر منظوری حاصل ہوئی۔ مشینی ترجے اور اُردو کوائفیہ کے سافٹ ویئر تیار ہوئے۔ کمپیوٹری لسانیات پر دو ورکشاپیں منعقد ہوئیں۔ مائیکروسافٹ کی کئی ملکی و غیر ملکی کانفرنسوں میں مجھے مقالے اور معروضات پیش کرنے کا موقع ملا۔ ان کاموں کا اثر اتنا ہوا کہ کمپیوٹر/ انٹرنیٹ/موبائل فون پر اُردو کا غلغلہ بلند ہو گیا۔ حکومتِ پاکستان نے بھی اپنے ویژن ۲۰۳۰ء میں اسے اہمیت دی۔ اس سارے عرصے میں کابینہ ڈویژن کے افسرِ صیغہ سید سرفراز شاہ مرحوم کی خدمات ہمیشہ یاد رکھی جائیں گی۔

۲۰۰۸ء میں اس کام کو انجام دیتے دس برس گزر چکے ہیں۔ اب وہ وقت آ گیا ہے کہ ایسی تمام نگارشات ، مقالات اور تحریریں ایک اہم علمی سرمایے کے طور پر جمع کر دی جائیں تاکہ اُردو اطلاعیات کا میدان اپنی دستاویزی بنیادوں پر بھی استوار ہو سکے اور مرکزِ فضیلت کے تیسرے مرحلے Institute for Urdu Informatics کا ان دس سالہ کاوشوں کے نتیجے میں آغاز ہو سکے۔ اس دوران میں یہ منصوبہ کئی نازک مرحلوں سے گزرتا رہا جن میں کامیابی سے ہمکناری پلاننگ کمیشن کے جناب طالب حسین ، خالد حسین سید اور کابینہ سیکرٹری سید مسعود عالم رضوی، ایڈیشنل سیکرٹری جناب اشتیاق احمد اور جائنٹ سیکرٹری محترمہ شمع خالد اکبر کی مرہون رہی۔

"اُردو اطلاعیات" کے نام سے کتابی صورت کا یہ سلسلہ جاری کیا گیا ہے۔ پہلی جلد میں میری اپنی

تحریریں شامل ہیں اور ان سے اگلی جلدیں میرے رفقائے کار اور دیگر اساتذہ اور طلبہ کی نگارشات پر مبنی ہیں جو وقتاً فوقتاً سامنے آتی رہی ہیں۔ اس مرحلے پر سیکرٹری کابینہ جناب غیاث الدین اور ایڈیشنل سیکرٹری جناب عطا محمد راجا قابلِ ستایش ہیں کہ اُن کی رہنمائی سے تیسرے مرحلے کی طرف اُٹھ رہے ہیں۔

یہاں میں اپنی ٹیم کے ان مختلف افراد کا ذاتی طور پر ممنون ہوں جو مختلف اوقات میں ساتھ دیتے اور میرے ساتھ کام کرتے رہے۔ ان سب کی مجموعی کوششیں آج آپ کے سامنے ہیں۔ ان کا ذکر مختلف تحریروں میں موجود ہے، اس لیے فرداً فرداً ان کا نام لینا مشکل ہے۔ ان میں سے بعض اپنا اپنا کردار ادا کر کے رخصت بھی ہوتے رہے لیکن اُردو اطلاعیات کا یہ ایک نیا میدان وضع کرنے میں خاطر خواہ حصہ دار بنے۔ اُردو اطلاعیات کی تاریخ انھیں ہمیشہ یاد رکھے گی۔

ڈاکٹر عطش دُرّانی
پراجیکٹ ڈائریکٹر
مرکزِ فضیلت برائے اُردو اطلاعیات

اسلام آباد، ۱۴۔ اگست ۲۰۰۸ء

۱

اُردو اطلاعیات: اُردو کا مستقبل

"اُردو اطلاعیات ہی اُردو کا مستقبل ہے"۔ یہ بات نوشتۂ دیوار ہے لیکن اسے سمجھنے اور جاننے کے لیے ہمیں چند لمحوں کے لیے اُردو کے مستقبل میں جھانکنا ہوگا۔ اکیسویں صدی کی ایک چوتھائی گزر جانے دیں تب ہمارے بچے کمپیوٹر کے سامنے بیٹھے اُردو تحریروں کا مطالعہ کر رہے ہوں گے، جملہ سازی سیکھ رہے ہوں گے، بامعانی نگارشات اور بے مقصد تحریروں کا فرق سمجھ رہے ہوں گے یا پھر چھوٹی چھوٹی کہانیاں اور مضمون لکھتے ہوں گے اور نظمیں تخلیق کرتے ہوں گے۔ ذرا بڑے بچے وہ ساری تحقیق سکول کی سطح پر انجام دیتے ہوں گے جو آج ایم اے کے طلبہ بڑے فخر کے ساتھ اپنے کارناموں کے طور پر پیش کرتے ہیں، جیسے "اشاریہ جات"، " کتابیات" اور "احوال و آثار" وغیرہ کی نام نہاد تنقیدی یا رائے زن تحقیق۔ اس دور میں سکولوں کے طلبہ آج کے جامعاتی طلبہ کو اسی طرح طفلِ مکتب سمجھیں گے جیسے آج کے طلبہ عکسی نقول کرنے والی مشینوں کی موجودگی میں ہاتھ میں نوٹس لینے والے کو بے مقصد وقت ضائع کرنے والے قرار دیتے ہیں یا پھر کیلکو لیٹر کی مہارت رکھنے والے لوگ پہاڑے یاد کرنے والوں اور گنتارے کی مشق کرنے والوں کو سمجھتے ہیں۔

کل کے کمپیوٹر/ ویب سائٹ/ انٹرنیٹ پر اُردو میں تحقیق کے بنیادی آلات، کتابیات سازی، لغات، اشاریے، تحقیقی جائزے، رپورٹیں، ادبی متون، اصول قواعد، مختلف ذخیرہ ہائے الفاظ، تحقیقی مسلیں/ فائلیں، پتے، عنوانات وغیرہ موجود ہوں گے اور اُردو کے محققین کو ان میں سر کھپانے اور کتب خانوں کی گرد جھاڑنے کی ضرورت بہت کم ہوگی۔ بنیادی کوائف اور ذخائر کمپیوٹر پر آ جانے سے کاغذی کتاب کا رواج بھی کم ہو جائے گا۔ سیاہی کے قلم کہ جگہ ماؤس، کلیدی تختے اور نوری قلم کو مل جائے گی۔ طلبہ کی کاپی کی جگہ کمپیوٹر نوٹ بک ہوگی اور کتاب/ رسالے کی جگہ ای بک اور انٹرنیٹ کو حاصل ہوگی۔ کمپیوٹر کی اگلی نسل آ جانے کے بعد تو شاید کمپیوٹر سکرین اور کلیدی تختہ بھی غائب ہو جائے گا اور اس کی جگہ برقیاتی چشمہ اور برقیاتی دستانہ لے لے گا۔ طلبہ انگلیاں نچائیں گے اور اپنے چشمے کے اندر ہی اندر پڑھیں لکھیں گے۔ یہ تو طے ہے کہ سب کچھ ہو رہا ہوگا، اگر اُردو میں نہیں تو پھر معاف کیجیے انگریزی میں تو ہو رہا ہوگا۔ اُردو تو کیا (خاکم بدہن) کوئی بھی پاکستانی زبان پھر کہیں بھی نہیں ہوگی۔ تب ایم اے اُردو یا پی ایچ ڈی اُردو کیا کرے گا؟ اس کی اعلیٰ سطحی مہارت کہاں درکار

ہوگی۔

ابھی وقت ہے۔ اگر آج ہم نے اُردو کے لیے مناسب تیاری کر لی تو پھر اس دور میں ہمیں ترجمہ کاری کے حوالے سے کم مائیگی سے بھی نجات مل چکی ہوگی۔ پوری دنیا اب انگریزی سے پیچھا چھڑانے کی فکر میں ہے۔ دن رات ترجمے کے سافٹ ویئر بن رہے ہیں۔ بس ایک کلید دبانے سے ترجمے کا سافٹ ویئر منٹوں میں انگریزی کی پوری کی پوری دستاویز، اقتباس یا کتاب اُردو میں پیش کر دے گا۔ اُردو کے اساتذہ کی تربیت کمپیوٹر ہی پر ہوگی اور طلبہ بھی کمپیوٹر پر تعلیم حاصل کر رہے ہوں گے۔ پابند جامعات کی جگہ شاید فاصلاتی تدریس لے لے گی۔

دوسری طرف ان تمام کاموں کی تیاری کے لیے ہمیں آج ہی سے کچھ کرنا ہوگا۔ اُردو کا مستقبل آج ہمارے ہاتھ میں ہے۔ یہی مستقبل اُردو کی نصاب سازی کا ہدف ہے، جس کی تیاری ہمیں آج کرنا ہے۔ ہمیں کمپیوٹر کے لیے اُردو متون کا انتخاب کرنا ہوگا۔ ان کی تدوین کرنا ہوگی۔ تحقیقی کوائف جمع کرنے ہوں گے۔ ان کے مناسب حصول کے لیے فارمولے بنانا ہوں گے۔ تدریسِ اُردو کے اسباق تیار کرنا ہوں گے۔ تکنیکی وتنقیدی اصول وضع کرنا ہوں گے اور ان کا تقابلی چارٹ بنانا ہوگا۔ کتب خانوں کی مسل داری کرنا ہوگی اور اشاریے وضع کرنا ہوں گے۔ اُصولِ ترجمہ اور اُصولِ اصطلاحات سازی کو وجود میں لانا ہوگا۔ تخلیقی امور میں معاونت کے اصول اپنانا ہوں گے۔ مطالعاتی مواد کی تشکیل کرنا ہوگی۔ کہانی لکھنے کی تکنیک بوجھو تو جانے اور ذخیرہ ءالفاظ کی معیاری فہرستیں وضع کرنا ہوں گی۔ تحریر کے اصول سامنے لانا ہوں گے۔ کمپیوٹر کی گرامر لکھنا ہوگی تاکہ اُردو ضابطہ برائے ترجمہ وضع ہو سکے اور انگریزی سے اُردو ترجمے کا سافٹ ویئر وجود میں آ سکے۔

ان سب باتوں کے لیے کمپیوٹر کی کون کون سے تکنیکیں وجود میں لائی جا سکتی ہیں۔ ذرا ایک نظر ان پر ڈال لی جائے۔

۱۔ سب سے پہلے اُردو کے مشینی قواعد یعنی ایسی گرامر جو شاخ داری (Branching) کے عمل سے گزار کر الگ الگ ضابطے یا کوڈ نمبر وجود میں لانے کا سبب بنے - ان کے مقابلے اور سیاق وسباق کے حوالے سے انگریزی گرامر کے کوڈ وضع ہوں گے اور کچھ ایسے طریقوں سے انگریزی - اُردو - انگریزی ترجمے کا سافٹ ویئر وجود میں آ جائے گا - اگر ماہرین زبان وادب نے اُردو کی اس گرامر کی تشکیل میں دلچسپی نہ لی تو یہ خواب شرمندۂ تعبیر نہ ہوگا۔ احمد مرزا جمیل جیسے خطاط اور نوری نستعلیق کے عارضی طریقے سے اُردو کی کمپوزنگ کا سافٹ ویئر تیار کروایا ہے جو اب ای میل اور انٹرنیٹ کی راہ میں حائل ہو رہا ہے-

۲۔ **اُردو ویب سائٹ:** انٹرنیٹ/سرچ انجن/ای ایڈریس وغیرہ ایسے امور ہیں جن میں براہ راست اُردو زبان کے ماہر کی مداخلت درکار ہے- ان کا ذکر آگے آئے گا لیکن فی الوقت اُردو کے نام پر جو

ویب سائٹس موجود ہیں وہ "تصویری صفحہ "پرعمل کرتی ہیں-ان میں تحقیق اور تلاش ممکن نہیں اور یہ کمپیوٹر پر یادداشتی جگہ (Memory) گھیرتی ہیں-

3- **اُردو کی ترجمہ مشین:** یہ سافٹ ویئر ابھی قواعد کے علاوہ ماہر زبان کی اور بھی مہارتوں کا طلب گار ہے؛ جیسے تعبیر مفہوم اور سیاق الفاظ کی تکنیکی پہلو، معیاری ذخیرہ الفاظ، طوالت الفاظ، کلیدی الفاظ(Key Words)، عمل تقلب(Transformation)، صرف ونحو کے بلاک وغیرہ-

مشینی ترجے (MT) کے کئی درجے اور سطحیں ہیں- ہر ایک کے لیے الگ لسانی تحقیق درکار ہے چونکہ CAHT (Comptur Assisted Human Translations) انسانی ترجمہ کاری ہے اس لیے اس میں نسبتاً کم ضرورت پڑے گی لیکن کمپیوٹر ترجمہ CAT(Computer Assisted) یا مشینی ترجمہ MAT(Machine Assisted) میں زیادہ تحقیق درکار ہے- دراصل یہ انسانی نہیں بلکہ مشینی کارگزاری ہوگی اور مشین کو ہر بات کے لیے ایک پروگرام بنانا چاہیے-ان کاموں کے لیے کتنے افراد اور اذہان کی محنت درکار ہے؟ اس کا اندازہ اس بات سے کر لیجیے کہ ۱۹۸۷ء میں جاپان میں مشینی ترجمہ کے منصوبوں پر نو سو افراد کام کر رہے تھے-

زبان دانوں کی درکار مدد والے اس کام میں جو مسائل کے حل میں درپیش ہیں، ان کا اندازہ کچھ یوں ہے:-

(۱) کثیر لسانی لغات درکار ہوں گے جن میں ہر لفظ اور جملے کا مکمل اور صحیح متبادل درج ہو-

(۲) کثیر المعانی الفاظ کی ہر رنگت کے مکمل اور صحیح متبادلات موجود ہوں-معنویات کے لحاظ سے یہ مشینی ترجے کا مشکل ترین مرحلہ ہے-

(۳) ایسے الفاظ جن کا ترجمہ دوسری زبانوں میں موجود نہ ہو جیسے اُردو کے " گوالا" اور "غیرت" اور انگریزی کے "موٹروے"، " کمیونٹی" اور " سول سوسائٹی" وغیرہ-

(۴) ایسے افعال اور زمانے جو دوسری زبانوں میں نہ پائے جاتے ہوں جیسے اُردو کے فعل مستقبل جاریہ اور ماضی شکیہ "وہ جا تا رہا ہوگا"-

(۵) کسی جملے کے اندر ترتیب الفاظ یا نحوی ترکیب کا تقابل- دوسری زبان کی نحوی ترکیب کے ساتھ ان الفاظ کا تعین جو کبھی فعل، کبھی اسم اور کبھی صفت کے طور پر استعمال ہوں، جیسے "بوڑھا"-

(۶) ترجے کی تکنیک کا اختلاف، برعکس ترجے کی صورت میں جیسے انگریزی سے اُردو کی تکنیک، اُردو سے انگریزی کی تکنیک سے مختلف ہوگی- نیز لفظ بہ لفظ ترجے اور بامحاورہ ترجے کی تکنیک بھی ایک دوسرے سے مختلف ہوگی-

(۷) ترجے کی صحت کی شرح ہدفی زبانوں (Target Languages) کے بدل جانے سے مختلف ہو

جاتی ہے جیسے سسٹران (Systran) کے نظام پر مشینی ترجمے کے سافٹ ویئر میں انگریزی سے فرانسیسی ترجمہ ۹۵ فیصد صحیح ہوتا ہے لیکن انگریزی سے اطالوی ترجمہ ۸۵ فیصد صحیح ہوتا ہے- ان کی انسانی ذرائع سے تصحیح کے اصول وضع کرنا بھی ماہرین اُردولسانیات ہی کا کام ہوگا-

(۸) اُردو کی برقیاتی کتب/ ڈسک میں متن اور اس کی خصوصیات، محضری تجزیے (Discourse Analysis)، عمومی اور خصوصی زبان، تھیسارس اور اس کے رابطے، لغات نگاری اور مابعد لغات نگاری ہیئتیں (Lexicographical and post lexicographical forms) اشتقاقات اور تصریف (Inflation) وغیرہ کی تکنیکی مسائل بندی وغیرہ-

(۹) کمپیوٹر خواندگی (Computer Literacy) کی تعریف، حدود، محضر اور اس کے ذخیرۂ الفاظ اور اصطلاحات کی تشکیل-

آیئے اب ان تحقیقی امور کا جائزہ لیتے ہیں جو اس سلسلے میں اُردو زبان میں انجام دیے جانے ہوں گے۔

(۱) حروف سے متعلق: اُردو حروف تہجی کیا اور کتنے ہیں؟ ان کی ترتیب مع حرکات اور اشکال، جوڑ- ان کے بنیادی اصول کیا ہیں؟ تعدد حروف (Letter Frequency) یعنی مختلف مضامین میں حروف وارد ہونے کی شرح کیا ہے؟ یہ تحقیق متعدد وقفوں سے درکار ہوتی ہے- املاء اور تدریسی اصول، نقل حرفی (Transliteration) کے اصول جو موضوعی (Subjective) سے ہٹ کر رجوعی (Reversible) ہوں اور کمپیوٹر کی ضروریات پوری کرسکیں- اُردو میں نقل حرفی کے لیے موجود 35 سکیمیں کمپیوٹر کے لیے ناکافی رہی ہیں- رسم الخط کی بحث، ترسیمے (Ligature) کے اصول، نسخ اور نستعلیق کا علمی مطالعہ، تکنیکی تجزیہ، جوڑ، کسی فانٹ کی اقسام کے اسباب وغیرہ-

(۲) الفاظ سے متعلق: الفاظ کا مطالعہ، ترسیمے کی ضرورت، معیاری املا، اگر ایک سے زیادہ معیارات ہوں تو ان کی جدولوں کی تشکیل- اُردو الفاظ کی تکمیلی طوالت ہجوں کی تحقیق اور اعراب حرکات، رموز اوقاف وغیرہ -کمپیوٹری اصول کی تشکیل، قواعدی صورتوں کا تعلق اور نسبت وغیرہ- متعدد الفاظ کا مطالعہ، جیسا کہ بارکر نے کینیڈا میں کیا تھا، ویسا مطالعہ مختلف محضروں کے حوالے سے، جیسا برمنگھم یونیورسٹی نے کو بلڈ (Co-build) ڈکشنری نے الفاظ کے عمومی اور خصوصی تعددِ استعمال معلوم کرنے کے لیے کیا تھا-

(۳) جُملے سے متعلق: اُردو جملے کی ساخت، کمپیوٹری نحوی ترکیب کی تشکیل، ٹیکسٹ ایڈیٹر کے اصول، اُردو جملوں کی تبدیلی (زمانوں اور حالتوں کی صورت میں) اور اس کی جدولیں، ان صورتوں کی وجوہ پر تحقیق، لفظ و معنی کا رشتہ، تھیسارس کے شجرے اور جدولیں، جملوں میں کلیدی، غیر کلیدی، مصدری/فعلی اور غیر مصدری فعلی الفاظ کا تعین، ذخیرہ ءالفاظ کی نامیاتی تشکیل، نتائجی الفاظ (Pragmatic Vocabulary) وغیرہ-

(۴) تحریر سے متعلق: متن اور متنیت، توسل (Mediation)، ابلاغ، محضر، اسلوب، طرز بیان اور لسانی سانچوں (Registers) پر بنیادی تحقیق، تحریر اور متن کے نفسیاتی عوامل، مطالعاتی مواد کی تشکیل، لسانی فنون، انشائی مسل داری، امکانات وغیرہ۔

(۵) تدریسِ اُردو: معیاری متون کی تشکیل، تدریسی اسباق، سٹوری مشین (پلاٹ کے مختلف امکانات اور کہانی آگے بڑھانے کے طریقے) الفاظ کے بُنت کار (Word Weavers)، نظم، آہنگ، عروض کے سافٹ ویئر کمپیوٹر نظم (Compupoem) سپرنگ بورڈ اور مکالمات کے سافٹ ویئر (تدریس ادب میں معاون)، زبان و بیان کی ذیلی معمولات (Subroutines)، انشاء پردازی کے تکنیکی اصول، سہ پیرا اور پنج پیرا انشائیہ (Three and Five Paragraph Essay)، اختلاف انتقاد ادب کی وجوہ اور امکانات کے سافٹ ویئر نیز دیگر تعلیمی سافٹ ویئر وغیرہ۔

(۶) اعلیٰ تحقیق، اُردو سے متعلق: ایسے تمام امور میں مزید کئی پہلوؤں پر تحقیق کرنے والوں کو مدد دینے کے لیے مختلف کتب خانوں میں موجود اشاریوں، متون اور معلومات کی کوائف کاری اور عملی ضروریات۔

مستقبل کے ایسے کئی امکانات ہیں مثلاً الیکٹرانک بک میں متون کی اشاعت جو ادیبوں اور شاعروں کو ورطۂ حیرت میں ڈالے ہوئے ہے اور اُردو کے طالب علم کو بھی اس نئی دنیا سے واسطہ پڑنے والا ہے۔

چنانچہ کم از کم ایم اے اُردو کے طالب علم کو "اُردو اطلاعیات" کے ایک پرچے میں مندرجہ ذیل حدود و قیود کا مطالعہ اور کمپیوٹر کا علمی استعمال آنا چاہیے۔

(۱) اُردو مشینی دور میں:

اُردو طباعت کا آغاز، ٹائپ اور رسم الخط کی بحث، پروفیسر محمد سلیم، ڈاکٹر طارق عزیز اور مقتدرہ کی شائع کردہ ایسی کئی کتابیں زیر مطالعہ لائی جاسکتی ہیں۔

(۲) اُردو لفظ کار کے امور:

اُردو لفظ کار یا ورڈ پراسیسر کی تاریخ ۔ ۱۹۸۰ء میں احمد مرزا جمیل کے ترسیمہ بنیاد (Ligature Base) پر تیار ہونے والا نوری نستعلیق۔ اس کے مسائل اور حل کی کوششیں، لائنو ٹائپ، شاہکار اور ان پیج کا سافٹ ویئر، ان کے استعمالات پر جہانگیر بک ڈپو لاہور کی شائع کردہ کتابیں وغیرہ۔

(۳) اُردو حرف کار کے امور:

اُردو میں حرفی بنیاد (Character Base) والے کمپیوٹر کی ضرورت، مقتدرہ کی ذیلی کمیٹی برائے یونی کوڈ اور اس پر حالی حروف اور نقطوں کے حوالے سے کی گئی کوششیں ۔ فاسٹ لاہور ۱۲۔ ستمبر ۱۹۹۸ء اور مقتدرہ قومی زبان کے سیمینار ۵۔ جون ۱۹۹۹ء نیز ۱۳۔ اپریل ۲۰۰۱ء کی رودادیں۔ احمد مرزا جمیل کے بیس سال

بعد ۲۰۰۰ء میں اُردو حروف پر مبنی ضابطہ تختیوں (Code- Plates)، کلیدی تختے (Keyboards)، ترتیب حروف (Collation Order)، نقل حرفی (Transliteration) جیسے پہلوؤں پر مقتدرہ قومی زبان، نادرا، فاسٹ لاہور، بین الاقوامی اسلامی یونیورسٹی، اسلام آباد اور عبادت فاؤنڈیشن وغیرہ کی کوششیں۔ Ghost Character Theoryk کا مطالعہ، "اخبار اُردو" کے مضامین، کتاب "پاکستانی اُردو: مزید مباحث" اور "اُردو ضابطہ تختی" وغیرہ کا مطالعہ۔ ان میں بیان کردہ موضوعات پر مزید مطالعے کی راہیں اور تحقیق۔ تعدد حروف پر مرحوم ڈاکٹر محمد افضل (۱۹۵۸ء)، ڈاکٹر ممتاز منگلوری ٹیکسٹ بک بورڈ پشاور (۱۹۸۲ء)، نادرا اسلام آباد (۱۹۹۹ء) کی تحقیق کے مطالعہ اور کلیدی بورڈ، ڈیٹا بیس کی تیاری کے لسانی مسائل، اُردو ای میل اور انٹرنیٹ کی تیاری پر پی ڈی ایم ایس (PDMS) کراچی، راقم، سافٹ ویئر لاہور وغیرہ کے کام۔ فاسٹ لاہور کے تیار کردہ "نفیس نستعلیق" اور مرکزِ فضیلت کے "پاک نستعلیق" فانٹ کی کارکردگی۔

(۴) اُردو سافٹ ویئر کے امور:

اُردو سافٹ ویئر بنانے کے موضع پر پہلا مقابلہ مقتدرہ قومی زبان نے 13- اپریل 2001ء کو منعقد کرایا۔ اس سلسلے میں 22 تا 29 مارچ 2001ء کو کراچی آئی ٹی سی این ایشیاء (ITCN Asia) کی نمائش منعقد ہوئی۔ اس کی روداد یں "اخبار اُردو میں شائع ہوئیں۔ اُردو سافٹ ویئر کے سلسلے میں انفارمیشن ٹیکنالوجی ڈویژن کے اعلانات، ایس ڈی این پی (SDNP) اسلام آباد اور فاسٹ لاہور کی کوششوں کا مطالعہ۔ مرکزِ فضیلت کی کوششیں قابل ذکر ہیں۔

اُردو سافٹ ویئر کی تیاری کے سلسلے میں مختلف اداروں کے تجربات خاص طور پر جعفر برادرز، نادرا، پی ڈی ایم سی، کوئیک سافٹ، انعام علوی نیز ایس ٹی ایس وغیرہ۔ مائیکروسافٹ اور انک سافٹ کی مقامی کاری (Localization) کو ملحوظ رکھنا ہوگا۔

Sustainable Development Network Programme (SDNP) اسلام آباد کے ساتھ ملٹی لینگوئیل انٹرنیٹ (MINC) اور مقتدرہ کے تعاون سے 28- جولائی 2001ء کو منعقدہ سیمینار برائے انٹرنیٹ وغیرہ کا جائزہ ''اخبار اُردو'' اسلام آباد کے کمپیوٹر نمبر، اُردو سافٹ ویئر نمبر (2002ء) اور مائیکروسافٹ آفس اور ونڈوز، وسٹا جیسے سافٹ ویئر اور موبائل فون وغیرہ اُردو میں منتقل کرنے کے تجربات اور معیاری برقیاتی فرہنگ کی مقتدرہ کی طرف سے اشاعت اس موضوع پر مطالعاتی مواد مہیا کریں گے۔

(۵) مشینی ترجمے کے امور:

انگریزی اُردو ترجمے کے سافٹ ویئر کی تیاری کے سلسلے میں یونیورسٹی گرانٹس کمیشن اسلام آباد اور پشاور یونیورسٹی کے زیر اہتمام ۱۰ تا ۱۳ اگست ۱۹۹۱ء کو باڑہ گلی میں منعقد ہونے والے سیمینار کی روداد۔ فاسٹ لاہور اور مقتدرہ قومی زبان کی کوششوں اور مطبوعات کا مطالعہ۔ خاص طور پر پشاور یونیورسٹی کے پروفیسر ایم اے

خان کی کتاب"Text based Machine Translations" اورڈاکٹر عابد خان (پشاور یونیورسٹی) کے مقالات قابل ذکر ہیں- علاوہ ازیں باڑہ گلی میں پشاور یونیورسٹی کے قومی سیمینار میں ڈاکٹر محمد عابد خان، محمد قاسم، ارشد شیخ اور عظمیٰ یوسف وغیرہ کے مقالات اُردو کی کمپیوٹر گرامر کے مسائل وغیرہ سے متعلق ہیں- مثال کے طور پر ان میں سادہ اُردو جملے، اسمیہ جملے، فعلی جملے، صفاتی جملے وغیرہ کمپیوٹر ضابطوں کے حوالے سے دیکھے گئے ہیں- ان کی اصولی ہیئت (Canonical Form) کا انگریزی کی اصولی ہیئت سے مقابلہ کیا گیا ہے- ان کے عناصر (Factors) کا تجزیہ کیا گیا- اگست ۲۰۰۷ء میں پشاور یونیورسٹی کی Language and Technology 07 کانفرنس اور جنوری ۲۰۰۹ء میں فاسٹ میں CLT09 کانفرنس کی روداد یں۔

کمپیوٹر لٹریسی کے حوالے سے اصطلاحات سازی کی جارہی ہے اور اس کے اسالیب ومحضر پر بھی تحقیق انجام دی جارہی ہے جوای میل اور انٹرنیٹ کے الفاظ میں وضع ہو رہے ہیں- عربی میں یہ پہلے ہی وضع ہو چکے ہیں- ان کا مطالعہ اور آئندہ کے لیے تحقیق کی محفوظ راہیں تلاش کرنا ضروری ہے- کمپیوٹر کی اصطلاحات پر مقتدرہ نے بھی ایک ڈکشنری شائع کی ہے- اس سلسلے میں کئی مضامین "اخبار اُردو" کی زینت بنے ہیں- علاوہ ازیں کمپیوٹر پر اُردو کے حوالے سے تحقیق اور مباحث کا ایک بہت بڑا ذخیرہ مقتدرہ کے اندر وجود میں آ رہا ہے جو روز افزوں ہے- ابتدائی تعارف کے لیے مقتدرہ نے "اُردو اطلاعیات" کے نام سے ایک کتاب بھی شائع کی ہے- اُردو میں تکنیکی اور سائنسی موضوعات پر "ٹیکنیکل سائنس" اور " گلوبل سائنس" وغیرہ کے نام سے کئی رسالے شائع ہوتے ہیں- اُردو میں اُردو سائنس بورڈ لاہور کے علاوہ کئی ناشروں نے بھی کمپیوٹر ٹیکنالوجی پر کتابیں شائع کی ہیں- ان کا ذکر اور مطالعہ کیا جاسکتا ہے-

کمپیوٹر اور زبانوں کے سلسلے میں عالمی معیارات آئی ایس او(ISO)، یونی کوڈ(UNICODE)، مائیکروسافٹ(Microsoft) اور OXML وغیرہ کے ناموں سے پہچانے جاتے ہیں-

مقتدرہ نے یونی کوڈ کی رکنیت حاصل کر رکھی ہے تا کہ اُردو کو عالمی معیارات میں جگہ دی جاسکے- یہ کام ممکن ہے زبان کے ماہرین کی مدد کے بغیر مکمل ہو جائے لیکن نوری نستعلیق کی طرح اُردو کی ترقی اور مستقبل کی تمام تر ضروریات کا احاطہ شاید نہ کر سکے- یونی کوڈ میں حروف کی ساخت، املاء، خط، شکل اور معنویات کے حوالے سے کئی مباحث درکار ہیں جو وجود میں آ رہے ہیں- ان سے کمپیوٹری ادبیات کا بھی ایک بڑا ذخیرہ تشکیل پا رہا ہے- ان اداروں میں کی گئی کوششیں۔

ضرورت اس امر کی ہے کہ ان بنیادی خاکوں کے حوالے سے جامعاتی سطح پر تحقیق، تدریس اور تعلیم انجام پائے- جامعہ اسلامیہ بہاولپور نے یہ کام شروع کر دیا ہے- شعبہء اُردو جامعہ پشاور کا سیمینار (۲۶ تا ۲۸ اگست ۲۰۰۱ء) " تدریسِ اُردو اور نصاب " اس امر کا شاہد ہے کہ اس سے اُردو کے نصاب کی تشکیلِ نو میں خاطر خواہ مدد ملے گی- اس کی رپورٹ اخبار اُردو کے شمارہ اکتوبر ۲۰۰۱ء میں شائع ہو چکی ہے- جامعہ سندھ کے

ارباب بست وکشاد پشاور یونیورسٹی کے ان نتائج اور مقتدرہ کی کاوشوں کے منتظر ہیں تاکہ وہ بھی اپنے نصابات میں اسے شامل کرسکیں اور اکیسویں صدی کے معاصر بن سکیں-IBA کراچی کے جشن طلائی کے موقع پر مائیکروسافٹ کا اُردو میں نصاب سازی اور تدریس کا اعلان اس سمت میں ایک واضح قدم ہے۔ مائیکروسافٹ اور ہائر ایجوکیشن کمیشن کے اقدامات بھی قابل توجہ ہیں۔

اس مرحلے پر اہم سوال یہ ہے کہ اُردو کے نصاب میں درس وتدریس کے حوالے سے اب کون کون سے اقدامات کیے جاسکتے ہیں۔

(الف) پرائمری سطح پر دیکھیں تو اُردو کا کوئی معیاری قاعدہ سرے سے وجود ہی نہیں رکھتا- اُردو کے حروف تہجی، حروفِ املا اور عمل تارید کے الگ الگ ہونے کے باوجود کسی بھی قاعدے میں حروف کی تعداد یکساں نہیں-کوئی قاعدہ حروف تہجی کی بنیاد پر ہے اور کوئی حروف املا کی بنیاد پر-حروف کے جوڑ اور ان کی نشستوں کے لیے "خالی حروف" (Ghost Characters) کے حوالے سے جو تحقیق سامنے آئی ہے، اسے پرائمری سطح پر اُردو لکھنے اور پڑھنے کے لیے استعمال کیا جاسکتا ہے- ابھی صرف خواندگی کے کچھ قاعدے (پنجاب، سرحد وغیرہ میں) یوں وجود میں آئے ہیں۔

پرائمری کے نصاب میں متعدد حروف اور متعدد الفاظ کی بنا پر درجہ وار ذخیرۂ الفاظ مرتب کرنا اور ابتدائی سطح پر اُردو کے بنیادی اور ثانوی الفاظ کے امتیاز کی بنا پر بنیادی الفاظ اور سادہ جملوں کی تشکیل، صرف ونحو کے جدولوں کے حوالے سے نصابی کتب کی تدوین وتشکیل اور کمپیوٹر لٹریسی کا پرائمری سطح پر آغاز، نیز کمپیوٹر کے ذریعے قاعدے اور ابتدائی کتابوں کی تدریس شامل کرنا ضروری ہے-

(ب) ثانوی سطح پر مختلف محضروں (Discourses) کا تجزیہ، کیمبرج کے محضری تجزیے (Discourse Analysis) کی طرز پر کیا جانا اور پھر اس کے مطابق نصابی سرگرمیوں اور کتابوں کا وجود میں آنا لازم ہے-کمپیوٹر کے بارے میں آگاہی اور اُردو پراسیسر پر اساتذہ اور طلبہ کی عملی تربیت ضروری ہے- اُردو ای میل، انٹرنیٹ اور ویب سائٹ کا استعمال مضمون نگاری، انشاپردازی وغیرہ کے حوالے سے بھی ضروری ہے- اُردو کا کوئی لغت داخلی ترتیب حروف (Collation Order) کے لحاظ سے یکسانیت کا حامل نہیں جبکہ ثانوی سطح پر طلبہ کو اشاریہ سازی کی تربیت دی جانی چاہیے-

(ج) اعلیٰ ثانوی سطح پر ادبیاتی ذخیروں کے لیے کمپیوٹر انٹرنیٹ کے استعمال، اقتصادی اور دیگر معلومات، کوائف وغیرہ سے آگاہی، کمپیوٹر لٹریسی سے متعلق مواد کی نصابی سرگرمیوں اور کتب میں شمولیت لازم ہے-

پرائمری سے اعلیٰ ثانوی سطح تک اُردو کے نصاب کا نقشہ (Matrix) وضع کرنے والوں کو ایسے تمام امور کا ابھی سے لحاظ کرنا چاہیے تاکہ ان کے تیار کردہ نصاب سے تعلیم پانے والے بچے اگر آج سے بیس

پچیس سال بعد عملی زندگی میں آئیں تو وہ اُردو کمپیوٹر لٹریسی کے عہد سے مطابقت اختیار کر سکیں اور زندگی کی دوڑ میں پیچھے نہ رہ جائیں۔

(د) گریجویٹ اور پوسٹ گریجویٹ سطح پر ان تمام مذکورہ بالا امور کا مطالعہ اور ان سے متعلق مواد کا احاطہ کرنا چاہیے۔ فی الوقت بعض جامعات اُردو کمپیوٹر کی عملی تربیت کو بھی ایم اے اُردو میں شامل کیے ہوئے ہیں۔ یہ انڈر گریجویٹ سطح پر مکمل ہو جانی چاہیے۔ نیز ایم فل اور پی ایچ ڈی کی سطح پر ان پہلوؤں کے حوالے سے تحقیق انجام دی جانی چاہیے۔ خاص طور پر اشاریہ سازی، کتابیات وغیرہ کے علاوہ تکنیکی لسانی امور کو بھی ان کی مناسب حیثیت ملنی چاہیے۔

(۱) منصوبہ ساز مجلس کا قیام:

ایک کمیٹی یا مجلس کا قیام جو کمپیوٹر کے ماہرین، اُردو زبان کے ماہرین، تحقیق، نصاب سازی اور تعلیم کے ماہرین پر مشتمل ہو اور وہ سارے پروگرام کو وضع کرنے کی پابند ہو۔

(۲) نصابی منصوبے کی تشکیل:

نصاب سازی چند ثانیوں، دنوں یا ہفتوں کا کام نہیں ہوتا۔ اس کے لیے ایک باقاعدہ منصوبہ بنایا جاتا ہے۔ منصوبے کے مطابق پائلٹ جانچ کی جاتی ہے اور یوں اس کے نفاذ کے لیے کئی تعلیمی سال درکار ہوتے ہیں چنانچہ ایسا ایک طویل المیعاد منصوبہ تشکیل دیا جائے۔

(۳) اہداف اور سطح کا تعین:

اہداف (Goals) اور سطح (Level) کے حوالے سے کمپیوٹر کی تعلیمی استعداد (Learning Competencies) کا تعین کرنا ضروری ہے تاکہ اس کی روشنی میں کمپیوٹر کا کرداری ماحصل (Behavior Outcome) معلوم کیا جا سکے۔

(۴) کمپیوٹری اطلاقات کا تعین:

کلیدی تختے، پروگرام کاری، لفظ کاری (ورڈ پراسیسنگ)، کوائف کاری (ڈیٹا پراسیسنگ) اور کوائفیہ (Database) کا جائزہ، نیز مختلف پروگراموں کے اطلاقات، استعمالات اور تعاملات (Functions) کا اُردو کے حوالے سے تعین لازم ہے۔

(۵) تدریسی معاونات کا تعین:

کمپیوٹر لٹریسی کے حوالے سے تدریسی معاونات کی ضرورت، اہمیت، وسعت، رہنمائی، تیاری وغیرہ تعلیم اور کمپیوٹر کے ماہرین کے ساتھ مل کر ہی انجام دی جا سکتی ہے چنانچہ ان معاونات کا تعین کر لینا لازم ہے۔

(۶) اساتذہ اور منتظمین کی تربیت:

اُردو کمپیوٹر کے استعمال سے متعلق سکول کے اساتذہ کے لیے محکمہ تعلیم میں اور جامعاتی اساتذہ کے لیے اعلیٰ تعلیم کے اداروں (HEC) وغیرہ میں تعارفی اور تربیتی کورسوں وغیرہ کی تیاری اور نفاذ لازمی حیثیت رکھتا ہے۔

(۷) اُردو سافٹ ویئر کمپنیوں سے رابطہ:

ہر جامعہ اور تعلیمی ادارے کے لیے لازم ہونا چاہیے کہ وہ اپنے تدریسی پروگراموں کے سلسلے میں اُردو سافٹ ویئر کمپنیوں سے رابطہ کرے اور تدریسی ڈسکیں تیار کرائے۔ اب ان پر کوئی زیادہ لاگت بھی نہیں آتی۔

(۸) اضافی اُمور:

اُردو اطلاعیات کے مضمون میں کمپیوٹر کا تعارف، تاریخ، پروگرام، زبان، لوگرتھم، اطلاقات، ڈیزائن، گراف، سپریڈشیٹ اور ورڈ پراسیسنگ کے تکنیکی اموراور اُردو اطلاعیات کے تحقیقی اُمور کو بھی شامل کیا جا سکتا ہے۔ قلم کاری کی طرح اب کلیدی تختے کی مہارت کو بھی ان اضافی امور میں شامل کر لیا جائے تو اُردو اطلاعیات کی تدریس کے مقاصد بخوبی حاصل ہو سکتے ہیں۔

اس مطالعے اور ان حوالوں سے "اُردو اطلاعیات" کی تدریس کا خاکہ کچھ یوں بنتا ہے۔

۱۔ کمپیوٹر کا تعارف جس میں اُردو پروگراموں کا تعارف بھی شامل ہوگا، نیز کمپیوٹر کے موضوع پر شائع کیے گئے اُردو لٹریچر کا تعارف۔

۲۔ کمپیوٹر پر اُردو کے مسائل، معیار بندی، تحقیق اور مسائل۔

۳۔ اُردو املاء، ٹائپ اور دیگر تکنیکی امور، ان کی بناوٹ، تعدد استعمال وغیرہ۔

۴۔ اُردو صوتیات اور حروف، الفاظ اور معنی کا رشتہ، تعبیر اور سیاق وسباق، بنیادی لفظیات، ثانوی الفاظ، مترادفات ومتضادات پر تھیسارس کی بنیادیں، الفاظ کا نامیاتی رشتہ وغیرہ۔

۵۔ اُردو ترجمہ کاری اور اُصول لغات نگاری کی تکنیک، کمپیوٹر گرامر اور لسانیات وغیرہ کے امور۔

۶۔ تحریر، متن کی ساخت، ابلاغ، توسّل، اسلوب، طرزِ بیان وغیرہ کے مباحث۔

۷۔ اُردو کی لسانی اور ادبی، تکنیکی اور حوالہ جاتی ترقی کے امکانات اور مباحث۔

۸۔ لسانی تجربہ گاہ اور کمپیوٹر کے عملی استعمال کی مشق۔

ان خطوط پر مزید تحقیق بھی انجام دی جاسکتی ہے، غور وفکر ہوسکتا ہے۔ بہت کچھ اور بھی شامل کیا جا

سکتا ہے، لیکن ایک بات طے ہے کہ اُردو کی نصاب سازی اور تدریس میں اب کمپیوٹر کو دخل حاصل ہو چکا ہے اور اس کے لیے اُردو والوں کو اپنا بنیادی تحقیقی فریضہ انجام دینا ہوگا۔

(پشاور یونیورسٹی شعبہ اُردو کے سیمینار "تدریسِ اُردو اور نصاب"، منعقدہ ۲۶ یا ۲۸ اگست ۲۰۰۱ء، باڑہ گلی میں پڑھا گیا۔ "اخبارِ اُردو"، اسلام آباد، فروری ۲۰۰۲ء میں شائع ہوا اور جولائی ۲۰۰۸ء میں نظرِ ثانی کی گئی۔)

۲

اُردو صوتیے، نئ املا اور کمپیوٹر تختیاں
(جدید برقیاتی تقاضوں کے حوالے سے)

برقیاتی ڈاک (E.Mail) اور اطلاعیات کے عالمی جال (Internet) کے باعث دنیا بھر کی زبانوں کو تکنیکی حد بندیوں اور جبر سے واسطہ پڑ رہا ہے۔ چونکہ یہ ایجادیں اہلِ مغرب نے کی ہیں۔ اس لیے انھوں نے ہر قسم کی اختراعات میں اپنی زبانوں ہی کو پیشِ نظر رکھا اور تمام ٹیکنالوجی انھی کے حوالے سے وجود میں لائے۔ جب یہ ایجادات دوسری زبانوں تک پہنچیں، تب بھی معیارات اور ترکیبات مغربی زبانوں ہی کے حوالے سے سامنے لائے گئے۔ یہیں سے ان دوسری زبانوں کی مشکلات کا آغاز ہوا، جن میں ٹیکنالوجی پروان نہیں چڑھی۔

مقتدرہ میں اُردو کمپیوٹر کی ایسکی ضابطہ تختی (ASCII Code Plate) اور کلیدی تختے (Keyboard) کی معیار بندی اور اندرونی پیش کش/اظہار (Internal Representation) کے سلسلے میں یہ بات کھل کر سامنے آئی ہے کہ:

۱۔ ابھی تک اُردو کے حروفِ تہجی کی تعداد معین نہیں۔
۲۔ یہ تعداد روز افزوں ہے اور نئے صوتیے شامل ہو رہے ہیں۔
۳۔ ان حروف کے لیے نئ املا درکار ہے۔
۴۔ کمپیوٹر میں ان امور کی گنجایش رکھنا مقصود ہے۔

یہ نتائج اردو کے کسی نقص، تناقض یا کمی کو ظاہر نہیں کرتے بلکہ جدید لسانیات کے اصولوں کے عین مطابق ایک زندہ، متحرک اور فعال ترقی پذیر زبان ہونے کی شہادت فراہم کرتے ہیں۔ جبکہ انگریزی جیسی جدید، ترقی یافتہ زبان کا یہ پہلو بے حد کمزور اور غیر تسلی بخش ہے۔ اس کے حروفِ تہجی کی تعداد (۲۶) ان برقیاتی ایجادات سے صدیوں پہلے معین ہو چکنے کے بعد اب نئے تقاضوں کا ساتھ نہیں دے رہی۔ مجبوراً ایجادات اور تکنیکی ترقیوں کو انھی کے مطابق محدود کرنا پڑ رہا ہے۔ یا پھر لسانیاتی تقاضوں کے پیشِ نظر صوتیاتی ابجد (Phonetic Alphabets) وضع کرنا پڑ گئے ہیں اور یوں ایک نئ ابجد، جو ابھی تک صرف لغات میں موجود ہے، انگریزی کی موجودہ ابجد پر حاوی ہوتی جا رہی ہے۔ اس کے باوجود لسانی مسائل ہیں کہ بڑھتے

ہی چلے جا رہے ہیں۔

اُردو کا معاملہ اس سے مختلف ہے۔ایک تو مسئلہ یہ کہ اسے انگریزی کے حروف تہجی اور علامات کی حدود میں فٹ کرنے کی کوشش کی جا رہی ہے۔اور دوسرے یہ کہ مختلف املائی تحقیقات کی روشنی میں معلوم ہو رہا ہے کہ محولہ بالا مسائل موجود ہیں اور انھیں حل کرنے کی ضرورت ہے۔

اُردو میں بنیادی طور پر ۳۷ حروف تہجی تسلیم کیے جاتے تھے۔اور ابن مقلہ کی روایت کے تحت ہم شکل املا میں وضع کیے گئے تھے۔خاص طور پر ''ٹ، ڈ، ڑ، وغیرہ۔ بھ۔ پھ جیسی بھاری، حلقومی آوازوں، یا صوتیوں کو دو حرفی مجموعہ (ب+ ہ)، (ت+ ہ)، (پ+ ہ) قرار دیا گیا۔اس قدیم روایت کے مطابق حروف یا حرفی مجموعوں کی کل تعداد ۴۸ بنتی تھی، جو کچھ حسب ذیل تھی۔

ا،

ب، پ، ت، ٹ، ث،

ج، چ، ح، خ،

د، ڈ، ذ، ر، ڑ، ز، ژ

س، ش، ص، ض،

ط، ظ، ع، غ،

ف، ق، ک، گ، ل، م، ن، و، ہ، ء، ی، ے۔

بھ، پھ، تھ، ٹھ، جھ، چھ، دھ، ڈھ، ڑھ، کھ، گھ۔

ان کے علاوہ بھی کئی صوتیے، آوازیں یا حروف موجود تھے، جو بولنے، پڑھنے اور لکھنے میں تو آتے تھے لیکن انھیں تسلیم نہیں کیا جاتا تھا۔مثلاً:

آ۔ آم، آب، وغیرہ میں،

رھ۔ سرھانا وغیرہ میں

لھ۔ کولھو، دلھا وغیرہ میں

مھ۔ تمھیں، تمھارا

نھ۔ ننھا، پنھارن

ںھ۔ منھ، مینھ وغیرہ میں

وھ۔ وھیل، وھائیٹ وغیرہ میں

یھ۔ یھاں وغیرہ میں۔

بعض روایت پسند لوگ آ کو الف اور مد کا مجموعہ قرار دیتے اور باقی حروف کو شوشے دار طریقے یعنی ''ہہ'' سے لکھ دیتے جیسے سرہانا، کولہو، دلہا، تمہیں، تمہارا، جمہورا، منہ، میڈ وہیل، وہائٹ، یہاں۔ صرف ننھا یا ننھیال ان کی دسترس سے بچا ہوا تھا۔ اس کے علاوہ بھی کئی اصوات تھیں جو بولنے میں آتی تھیں۔ لیکن انھیں ہم لکھنے سے کتراتے رہے یا انھیں املائی مجبوری کے تحت لکھتے رہے۔ جیسے۔

اء۔ علماء، انبیاء

اٰ۔ جیسے: ماٰخذ وغیرہ

ۃ۔ جیسے: زکٰوۃ، صلٰوۃ وغیرہ

م۔ جیسے: آم، نام وغیرہ

ٹ۔ جیسے: کرشٹ، پاٹی میں

ں۔ جیسے انبیا، چنبیلی میں (نون میم کی آواز)

ں۔ جیسے: جنگ، رنگ، میں نگ کی آواز

وِ۔ جیسے: خورشید، خوان وغیرہ میں

وٗ۔ جیسے: خواب، خواہش، خواتین میں

ؤ۔ جیسے: گاؤں، کوئیں، ہندوؤں، وغیرہ میں

ؤ۔ جیسے جاؤ، لاؤ وغیرہ میں

ء۔ جیسے: ہمزہ اضافت میں

عربی کی بعض علامات اعراب جیسے زکٰوۃ، بذاتہٖ، مثلاً نسلٍ، عبدہٗ''، وغیرہ ان پر مستزاد ہیں۔ کمپیوٹر اور انٹرنیٹ کی کوڈ پلیٹ وضع کرتے ہوئے ایسے بہت سے مسائل سامنے آتے ہیں۔ ان مسائل میں ان حروف کی ترتیب (Sorting Order)، اشاریہ بندی اور تلاش (Search) بھی شامل ہے۔ اُردو کے نیم مصوتے زبر، زیر، پیش، جنھیں غلطی سے اعراب کی علامات سمجھا جاتا ہے، مسائل میں اضافہ کرتے ہیں۔ درحقیقت عربی زبان کی ترتیب کاری میں ان اعراب کی قطعاً ضرورت نہیں ہوتی لیکن اُردو کا یہ لازمی حصہ ہیں۔ مثلاً ہم اُردو الفاظ زبر، زیر، پیش کے بغیر بھی لکھتے ہیں اور ان کے ساتھ بھی۔ ہر دو قسم کے املا رائج ہونے کی بنا پر ترتیب کاری میں ہر دو کا خیال رکھنا پڑتا ہے۔ چنانچہ اُردو کا ہر حرف (چند استثنائی حروف کے علاوہ) چار قسم کے حروف کی شکل اختیار کر جاتا ہے۔

ا، اَ، اِ، اُ

یہ چاروں حروف ترتیب کاری، اشاریہ سازی یا لغوی ترتیب میں اس طرح سے آتے ہیں۔ مثلاً گھن، گَھن، گِھن، گُھن وغیرہ۔

لغت میں حسب ذیل ترتیب سے آئیں گے۔

گھن، گُھن، گَھن، گھُن

چندحروف کوچھوڑ کر باقی میں اس کی تعداد کی ضرب حروف تہجی کے لیے کوڈ پلیٹ میں خانوں کی تعداد کو بڑھا دیتی ہے۔ایک اندازے کے مطابق ان کے لیے ڈیڑھ سو خانے درکار ہیں اور مزید خانے اُردو کی روز افزوں ترقی کے لیے بھی چاہئیں۔

اگر ہم ایک نظر واپس ڈالیں تو اس بڑی تعداد کا جواز واضح ہو جاتا ہے۔اُردو تہذیبی سطح پر پروان چڑھنے والی ایک زبان ہے، جسے مختلف ماخذوں سے آنے والے حروف اور صوتیوں کے املا کے لیے طریق املا مستعار لے کر اس میں اپنی ضرورتوں کے مطابق تبدیلیاں کرنا پڑیں۔

اُردو کے چند حروف عربی ماخذ سے آئے۔ان کا املا اسی طرح رائج رہا جیسے عربی یا فونیقی میں تھا۔ یہی صورت فارسی حروف کی تھی۔جو پہلے عربی یا فونیقی رسم الخط میں ڈھل چکے تھے۔کچھ مقامی اصوات تھیں۔ (جیسے ٹ،ڈ،ڑ، بھ، پھ، ے وغیرہ)۔جنھیں ملتی جلتی املا کے لحاظ سے وضع کر لیا گیا۔البتہ ان کا املا ترقی پذیر رہا، پہلے ٹ،ڈ،ڑ،کو دو نقطے اور لکیر کے ساتھ تۃ، ۃ، ۃ، یا چار نقطوں کے ساتھ ت، ذ، ژ ظاہر کیا جاتا تھا۔ کتابت کی کششِ قلم نے انھیں 'ط' میں بدل ڈالا۔سندھی املا میں البتہ یہ اگلا قدم 'ط' نہیں اٹھایا گیا۔ بھ،کو پہلے 'بہہ' یا پہہ ،لکھا گیا۔پھر دو چشمی ھ مقرر کر لی گئی۔''ے'' کے لیے پہلے یائے معروف ''ی'' ہی مستعمل رہی۔ پھر اسے نصف کر لیا گیا جیسے ''ی'' پھر اسے نیچے کی طرف الٹا کھینچ دیا گیا اور ''ے'' کی موجودہ صورت بن گئی۔

مقامی اصوات کو غلطی سے ہندی حروف کہا گیا۔دراصل یہ مقامی صوتیے یا حروف تھے۔کچھ حروف انگریزی کے ذریعے وارد ہوئے جیسے Wh کی آواز ''وھ'' وھیل وغیرہ میں۔بعض لوگ اسے ابھی تک اُردو میں داخل نہیں سمجھتے۔اسی طرح ایک انگریزی حرف V اُردو میں داخل ہونے کو ہے۔عربی سندھی کی حد تک اس کا املا'' ڤ '' کی صورت میں دیکھنے میں آیا ہے۔اُردو میں ابھی تک ''وْ'' ہی سے کام چلایا جا رہا ہے۔''X'' کی آواز پشتو کے ''خین'' کی ہے۔جو ''خس'' یا ''خش'' کی آواز دیتا ہے۔اُردو میں اسے ''ک،گ،خ،س،ش سے ظاہر کر دیا جاتا ہے۔

ابھی تک اُردو میں حروفِ تہجی کی صورت وزارتِ تعلیم، قاعدوں اور بورڈوں کے حوالے سے کچھ یوں ہے۔ان کی تعداد ۵۳ ہے اور ان کی ترتیب حسب ذیل مانی گئی ہے:۔

ا،آ، ب، بھ، پ، پھ، ت، تھ، ٹ، ٹھ، ث، ج، جھ، چ، چھ، ح، خ، د، دھ، ڈ، ڈھ، ذ، ر، رھ، ڑ، ڑھ، ز، ژ، س، ش، ص، ض، ط، ظ، ع، غ، ف، ق، ک، کھ، گ، گھ، ل، لھ، م، مھ، ن، نھ، و، ہ، ء، ی، ے۔

اس ترتیب میں پانچ حروف: آ، رھ، لھ، مھ، نھ، کا اضافہ تسلیم کیا گیا ہے۔نون غنہ ''ں'' کو حرف تو

تسلیم کیا جاتا ہے لیکن ترتیب میں شامل نہیں کیا گیا۔ ان کے علاوہ بعض اہل علم وقلم، ماہرین، زبان اور لغات نویس ل، نؓھ، وھ، اور یھ، کو بھی اُردو کے حروف مانتے اور لکھتے چلے آئے ہیں۔ رشید حسن خان اور قدرت نقوی ''وھ'' کو مانتے اور لکھتے ہیں۔ سید قدرت نقوی، جوش ملیح آبادی، ڈاکٹر جمیل جالبی اور بہت سے دیگر اہل علم وقلم نؓھ کو ''منؓھ'' وغیرہ میں استعمال کرتے ہیں۔ سید قدرت نقوی ''وھ'' کو ''وھیل'' وغیرہ میں۔ شان الحق حقی ''ن''، کو ''کرشڻ' مانک وغیرہ کے ''ڻ'' کی جگہ استعمال کرتے ہیں۔

ڈاکٹر سہیل بخاری نے بہت سے دیگر حروف پر بھی بحث کی ہے، جن میں سب سے اہم ''نگ'' کی آواز ہے، جو انگریزی میں اب "n" کی بجائے "ŋ" سے ظاہر کی جاتی ہے۔ شان الحق حقی ''یھ'' کے صوتیے کو تسلیم نہیں کرتے جو ''یھاں'' جیسے الفاظ میں مستعمل ہے۔ اسی طرح وہ ''وھ'' کا صوتیہ بھی تسلیم نہیں کرتے۔ رشید حسن خان اردو میں ڻ کو ''نڑ'' کی آواز سمجھتے اور ''ن'' سے لکھنا جائز قرار دیتے ہیں۔ حقی صاحب نے اسے علیحدہ حرف تسلیم کیا ہے۔ اور اس کا نام ''ڑاں'' رکھا ہے۔ ہم اسے ''نون'' کے وزن پر ''ڑون'' کا نام دینا چاہتے ہیں۔ البتہ حقی صاحب ''نؓھ'' کا حرف علیحدہ نہیں سمجھتے اور ''منؓھ۔ اونؓھ وغیرہ کو ''رنگ۔ ڈھنگ، کا نون غنہ ہی سمجھتے ہیں، البتہ وہ ایک نون بشکل میم کو الگ صوتیہ قرار دیتے ہیں جیسے انبار، سنبل، چنبیلی وغیرہ میں۔

دراصل نون غنہ کی کئی صورتیں ہیں۔ جن میں سے کچھ واضح ہو چکی ہیں اور کچھ واضح ہو رہی ہیں۔ منک، جنگ، رنگ وغیرہ میں یہ ک اور گ کے ساتھ مل کر الگ آواز۔ صوت یا حرف ںٌ کی صورت میں واضح ہے۔ ڈاکٹر سہیل بخاری نے اس کے لیے '' ںٌ '' (نون کے درمیان دائرہ) کی املا کی سفارش کی ہے۔ منٌک، جنٌگ رنٌگ جیسے۔ ہاتھ سے لکھتے ہوئے ان دائروں کو نقطے کی صورت میں لکھنے کا مغالطہ ہوسکتا ہے۔ اس کے لیے مزید جزم کی علامت لکھ دی جائے۔ ''نْ'' جنْگ، رنْگ وغیرہ میں۔ اس حرف کے لیے ایک نام بھی درکار ہے۔ اسے ''ڑونگ'' قرار دیا جائے تو بہتر اور موزوں ہوگا۔

صوتیۃً '' ڻ '' کو اردو میں لکھنے سے اس کے ''ٹ'' میں بدلنے کا امکان موجود ہے۔ جیسے مانک، ماٹک، ہو جائے گا۔ پانی، پاٹی، بن جائے گا۔ چنانچہ اسے نون ہی پر اضافی ط کی صورت میں ن لکھا جا سکتا ہے۔ اور یوں مانک، پاٹی، کرشن ظاہر ہو سکتے ہیں۔ یعنی لکھے، پڑھے، بولے جا سکتے ہیں۔ خاص طور پر پاکستانی اُردو میں اس تلفظ، لہجے اور صوت کی ضرورتیں بے حد محسوس ہوتی ہیں جو دیگر پاکستانی زبانوں اور عوام کی ضرورتوں کی بنا پر سامنے آ رہی ہیں۔

جہاں تک نون کی ایسی آواز کا تعلق ہے، جو میم میں بدل جاتی ہے اور جسے ہم نون میمہ بھی کہہ سکتے ہیں۔ اس کے لیے یہ ضروری ہے کہ یا تو اسے ہم میم ''م'' ہی سے لکھیں یا پھر ایک نئی علامت ''ن'' مقرر کر لیں اور تمبورد تنبورہ یا تنبورہ لکھیں۔

یہیں میم کی ایک آواز غنہ کے ساتھ ظاہر ہوتی ہے اور دراصل یہ میم غنہ ہی ہے۔ اسے م کیساتھ نون

غنہ ڈال کر لکھنا چاہیے۔ جیسے آم، دوم، کوآنْم، دونْم، تمبْورہ لکھا جانا چاہیے۔ اگر نون غنہ کی علامت کے ہمزہ یا نون میں بدل جانے کا اندیشہ ہے تو پھر حرف م کے اوپر غنہ کی علامت دے دیں، جیسے آمْ، دومْ، دوتمبْورہ وغیرہ۔

کچھ صوتیے پنجابی، لاہوری وغیرہ سے اُردو بول چال میں داخل ہیں۔ ان میں ب اور پ، ت اور ٹ، ک اور گ، یا بھ اور پھ، تھ اور ٹھ، اور کھ اور گھ کی درمیانی اصوات ہیں جو گورمکھی میں ککّے کہلاتی ہیں، اب اُردو بول چال میں بھی عام ہیں۔ ان کے اظہار کی صورت بھی نکالنا ہوگی۔ جیسے ٻ، ٻھ، ٿ، ٿھ، ڳ، ڳھ وغیرہ۔ بالکل اسی طرح جیسے سرائیکی میں بھاری ب کی صوت کے لیے حرف ''ٻ'' وضع کیا گیا ہے۔ سرائیکی/ملتانی کی یہ صورت بھی اُردو پر اثر انداز ہو رہی ہے۔ اس کی مزید بھاری صَوت ''ٻھ'' بھی ہے۔

واو معدولہ بھی اُردو میں خاصا پریشان کن مسئلہ ہے۔ اگر اسے بولنا ہی نہیں تو املا سے نکال ہی دیں۔ ''اصطلاحاتِ پیشہ وراں'' میں ایسا ہی کیا گیا ہے۔ پروفیسر سید محمد سلیم نے تو بجا طور پر کہا ہے کہ خواب اور خواہش میں دراصل واو معدولہ نہیں بلکہ یہ ''ua'' کی آواز ہیں جیسے Khuahish, Khuab۔ وہ اسے واو پر دائرہ لگا کر ظاہر کرنا چاہتے ہیں۔ جیسے خوْاب، خوْاہش دراصل یہ واو مخلوط یا ملواں واو ہے۔ اسے جزم کے ساتھ خوْاب، خوْاہش لکھا جا سکتا ہے۔ واو کی ایک آواز خود اسی حرف کو لکھنے میں آتی ہے۔ یہ ''واؤ'' اور ''واءَ'' کی درمیانی صوت ہے۔ گاؤں، پاؤں، ہندوؤں، میں بھی ظاہر ہوتی ہے۔ اور اسے یا تو ہمزہ کے ساتھ یا اس کے بغیر لکھا جاتا ہے۔ اسے بھی دراصل غنہ کی آواز اور علامت کے ساتھ پیش کرنا چاہیے۔ ''واؤ، گاؤں، پاؤں، ہندووں، وغیرہ۔ وَ کی آواز، روءَ، سوءَاء میں بھی ظاہر ہوتی ہے۔

نھے غنہ کا صوتیہ نْھ جو منْھ یا مونْھ، اونْھ، مینْھ، بانْھ میں ظاہر ہوتا ہے۔ ابھی تک حرف اور اس کے نام کا تقاضا کر رہا ہے۔ اسے نھے (نْھ) کی طرح ''اینْھ'' کہا جا سکتا ہے۔

الف کی بھی کئی اصوات ہیں، جن میں سے ا، !۔ ء، اء، وغیرہ مستعمل ہیں جیسے علماء، سنا، ماٰخذ وغیرہ ہیں۔ لیکن اب یہ لکھنے میں نہیں آ رہے بلکہ مقتدرہ نے تو انھیں حذف کرنے کی سفارش کی ہے۔

ی کا ایک صوتیہ یائے مخلوط ہے جو کیْا، لیْا وغیرہ میں آتا ہے۔، اسے بھی جزم کے ساتھ لکھا جانا چاہیے۔

عربی کی تائے مدورہ ''ۃ'' ابھی تک ہمارے ہاں مستعمل ہے، اسے بھی اُردو حروفِ تہجی میں شامل کر لینا چاہیے۔ اور ''ہ'' کے بعد رکھ دینا چاہیے۔

ابھی اور نہ جانے کتنی اصوات ظاہر ہونے کی منتظر ہیں۔ حقیقت یہ ہے کہ اُردو کے حروف تہجی ترقی پذیر اور روز افزوں ہیں۔ ان کی کئی مکتوبی وغیرہ، املائی وغیرہ املائی اور حرفی وغیر حرفی صورتیں ہمارے سامنے ہیں۔ کچھ ابھی تک پردہء اخفا میں ہیں۔ ان میں سے مجوزہ اسّی ۸۰ صورتیں کچھ یوں ہیں۔

ا، اٗ، آ (اُ) ب، بھ، ب، بھ، ب، بھ، پ، پھ، ت، تھ، ٹ،
ٹھ، ٹ، ٹھ، ث، ج، جھ، چ، چھ، ح، خ، د، دھ، ڈ، ڈھ، ڈ، ڈھ،
ذ، ر، رھ، ڑ، ڑھ، ز، ژ، س، ش، ش، ص، ض، ط، ظ، ع، غ،
ف، ڤ، ق، ک، کھ، گ، گھ، گ، گھ، ل، لھ، مٓ، م، مھ، ں،
نھ، ںٔ، لٰٓ، لٗ، ن، نھ، ۏ، ۏ، ۏ، وٓ، و، وھ، ہ، ۃ، ء، ۂ، ئ، ی،
ے، یھ۔

اب ان صورتوں، علامتوں، صوتیوں یا حرفوں میں سے بعض کو چند حالتوں میں محض دوسری املائی حالت قرار دے بھی دیں تو بھی ہمارے پاس کم از کم ساٹھ حروف بچ رہتے ہیں۔ محض ٹائپ اور کمپوزنگ میں مختلف علامت قرار دیں تو بھی ۶۵ حروف بچ رہتے ہیں۔ ان میں تائے مدورہ ۃ شامل ہے۔ اُردو کے یہ ۶۵ حروف تہجی کچھ یوں ہیں۔

۲۔ا، آ

۱۱۔ب، بھ، پ، پھ، ت، تھ، ٹ، ٹھ، ث۔

۱۷۔ج، جھ، چ، چھ، ح، خ

۲۲۔د، دھ، ڈ، ڈھ، ذ

۲۸۔ر، رھ، ڑ، ڑھ، ز، ژ

۳۶۔س، ش، ص، ض، ط، ظ، ع، غ

۴۴۔ف، ق، ک، کھ، گ، گھ، ل، لھ

۵۳۔مٓ، م، مھ، ں، نھ، ن، ںٔ، لٰٓ، نھ

۵۷۔ۏ، وٓ، وٓ، و

۶۰۔ہ، ۃ، ء

۶۴۔ی، ئ، ئ، ے، یھ

جہاں تک لغات میں حروف کی تختیاں قائم کرنے کا تعلق ہے، ان میں صرف ایسے حروف چھوڑے جا سکتے ہیں، جن سے الفاظ شروع نہیں ہوتے۔ لیکن خیال رہے کہ رھ سے رھواس (رہنے کا عمل/کیفیت) رُھب (مشت)، ڑ سے ڑُنگا (بیل) گاڑی کی گھنٹی، ڑون (لٰٓ)، (نیا حرف تہجی)، ڑھ سے ڑھیڑی (رھیڑی)، ڑھیڑ (ڈھلوان) موجود ہیں۔ البتہ مٓ، ں، نھ، ن، ںٔ، وٓ، وٓ، لٰٓ، ئ، ئ مشینی ہیں۔

اب اُردو ٹائپ اور کمپیوٹر کی تختیوں کے تعلق سے دیکھیں تو ان تمام حروف کے لیے الگ جگہیں قائم کرنے کی ضرورت محسوس نہیں ہوتی۔ دوچشمی ''ھ'' رکھنے والی بھاری اصوات، غنہ، اور جزم کی علامت رکھنے والے حروف کے لیے علیحدہ خانوں اور کلیدوں کی ضرورت نہیں۔ یہی علامتیں کافی ہیں جو کلیدی تختے پر جوڑی جا سکتی ہیں اور ضابطہ پلیٹ (Code plate) پر ایک منفرد حرف کی صورت اختیار کر سکتی ہے۔ اس طرح صرف

مندرجہ ذیل بیالیس حروف یا علامتوں کی ضرورت باقی رہ جاتی ہے۔

ا،آ،ب،پ،ت،ٹ،ث،ج،چ،ح،خ،د،ڈ،ذ،ر،ڑ،ز،ژ،س،ش

ص،ض،ط،ظ،ع،غ،ف،ق،ک،گ،ل،م،ں،ن،و،ہ،ء،ی،ے،ھ،ۂ،ۓ

ان کے ساتھ اگر زبر، زیر، پیش کو نیم مصوتے یا نیم حروف علت (Semi Vowels) مان کر شامل کرلیں تو یہ تعداد ۴۵ ہو جاتی ہے۔ ہمزہ اضافت ''ٔ'' اور زیر اضافت (ِ) نشان اور ترتیب کاری میں نہیں آتے، اس لیے یہ عام حروف ہمزہ اور زیر کے علاوہ اور مستزاد قرار پائیں گے اور دیگر علامات x،!،؟،ۓ وغیرہ کے ساتھ شریک رہیں گے۔

چونکہ لوکارتم کے انداز پر حرفِ صحیح، حرفِ زبر، حرفِ زیر، اور حرف پیش الگ الگ ہوں گے، اس لیے، زبر، زیر، پیش کی علامات کے لیے الگ خانوں کی ضرورت درپیش نہیں ہوگی۔ اُردو میں مصوتے یا حروف علّت (Vowels) چار ہیں۔ ا، و، ی، ے، ان کے ساتھ جب زبر، زیر، پیش کے نیم مصوتے شامل ہوتے ہیں تو ان کی آواز نئے صوتیے میں بدل جاتی ہیں۔ انھیں حروفِ لین کہتے ہیں۔ چونکہ انھیں زبر، زیر، پیش کے ساتھ ظاہر کیا جا سکتا ہے، اس لیے انھیں الگ حرف شمار نہیں کیا گیا۔ لوکارتم کے اصول کے تحت کمپیوٹر کی ضابطہ تختی میں حروف یوں ہونگے۔

ا،اَ،اِ،اُ،

آ (باقی درکار نہیں)

اء(ۓ،ءِ)

ب،بَ۔بِ۔بُ

پ،پَ،پِ،پُ

ت،تَ،تِ،تُ

ٹ،ٹَ،ٹِ،ٹُ،

ث،ثَ،ثِ،ثُ،

ج،جَ،جِ،جُ،

چ،چَ،چِ،چُ

ح،حَ،حِ،حُ،

خ،خَ،خِ،خُ،

د،دَ،دِ،دُ،

ڈ،ڈَ،ڈِ،ڈُ،

ذ،ذَ،ذِ،ذُ،

ر،رَ،رِ،رُ

ڑ،ڑَ،ڑِ،ڑُ

ز،زَ،زِ،زُ،

ژ،ژَ،ژِ،ژُ،

س،سَ،سِ،سُ،

ش،شَ،شِ،شُ

ص،صَ،صِ،صُ،

ض،ضَ،ضِ،ضُ،

ط،طَ،طِ،طُ،

ظ،ظَ،ظِ،ظُ،

ع،عَ،عِ،عُ،

غ،غَ،غِ،غُ،

ف،فَ،فِ،فُ،

ق،قَ،قِ،قُ،

ک،کَ،کِ،کُ،

گ،گَ،گِ،گُ،

ل،لَ،لِ،لُ،

م،مَ،مِ،مُ،

ں،(باقی درکارنہیں)

ن،نَ،نِ،نُ،

و،وَ،وِ،وُ

ہ،ہَ،ہِ،ہُ،

ۃ،ۃَ،ۃِ،ۃُ،

ء،ءَ،ءِ،ءُ،

ی،یَ،یِ،یُ،

ے(باقی درکارنہیں)

ھ(ےً ےً)

ےٛ (ےً ےً)

ےٚ (ےً ےً)

ہمارے خیال میں اُردو اب تکنیکی گھاٹ چڑھ چکی ہے۔ اب اس کے صوتیے، حروف، املا، ترتیب، اُصولِ لسانیات، زبان کی رہنمائی، ہماری بول چال اور ادبیات کی روایت نہیں کرے گی بلکہ برقیاتی ٹیکنالوجی (اُردو کمپیوٹر، ای میل، انٹرنیٹ، اطلاعیات وغیرہ) کرے گی۔ اب زبان کی ترقی (Language Development) اور لسانی منصوبہ بندی (Language Planning) خواہ کارپس (Corpus) سطح پر ہو، لغاتیات (Lexicology)، یا اصطلاحیات (Terminology) کے حوالے سے ہو یا معجم نویسی (Thesaurii Building) یا قسر یہ لغت (Dicautom) ہو، جو جدید لسانی ترقی کے لوازم ہیں۔ ان سب کی رہنمائی شاہراہِ اطلاعیات (Information Highway) کی لوکارتم کرے گی۔

ہمیں اپنے مستقبل کی آئینہ بندی اپنے ذاتی خیالات اور موضوعی (Subjective) نظریات سے ہٹ کر جدید معروضیت (Objectivity) اور تکنیکیت (Technicality) کی بنیاد پر کرنا ہوگی۔ یہی زبانوں کی منزل ہے اور یہی اُردو کا ہدف۔

("اخبارِ اُردو"، اسلام آباد، مئی ۱۹۹۹ء)

۳

اُردو صوتیے اور ضابطہ تختی

اخبارِ اردو''مئی ۱۹۹۹ء میں شائع شدہ مضمون''اردو صوتیے، نئی املا اور کمپیوٹر تختیاں'' کی صدائے بازگشت مُلکِ عزیز کے علاوہ بھارت اور دیگر ملکوں سے بھی سنائی دی، جہاں اُردو کی ترقی اور نئی تشکیل سے وابستہ اہل فکر و دانش موجود تھے۔ چنانچہ ضروری محسوس ہوا کہ اس مضمون کو جو ابتداء میں محض ایک رائے کی حیثّیت رکھتا تھا، ۵ جون ۱۹۹۹ء کو مقتدرہ کے اردو کمپیوٹر کی ضابطہ تختی پر منعقدہ قومی کمپیوٹر سیمینار کی روشنی میں اسے باقاعدہ تکنیکی پہلوؤں سے دیکھا جائے۔(دیکھیے:ضمیمہ ا تا ۵)

پہلی بات جو کھل کر سامنے آئی وہ یہ تھی کہ اُردو ایک ترقی پذیر زبان ہے اور اس کے حروف تہجی روز افزوں ہیں۔ چنانچہ ان کا ہمیشہ خیال رکھنا چاہیے۔خواہ آج ہم کسی نئے صوتیے یا نئے حرف کو تسلیم کریں یا نہ کریں لیکن ان کے املا اور کل کے لیے ان کی گنجایش موجود ہونی چاہیے۔اُردو زبان کسی خاص طبقے یا گروہ کی جاگیر نہیں جو اسے اپنے خود ساختہ، قدیم اور دوراز کار معیارات میں محدود کردے۔۱۳ مئی ۱۹۹۹ء کو مقتدرہ کی تکنیکی کمیٹی کے اجلاس میں جب اس بنیادی اصول پر پرکھا گیا تو شاید اُردو حروفِ تہجی کی تعداد اس قدر بڑھ گئی کہ کمپیوٹر کی ضابطہ تختی (Urdu Code plate) اس کی متحمل نہ ہوسکتی تھی۔ٹھونس ٹھانس کے جو ایسکی ضابطہ تختی وجود میں آئی وہ بھی ابھی تنقید کا نشانہ بن رہی ہے۔ دیگر عمومی علامات کی بھرمار کے علاوہ بھی تین سو جگہیں درکار تھیں۔ مد، شد، دوچشمی ھ، زبر، زیر اور پیش نکال کر بھی ۴۵ خانے استعمال میں آتے تھے۔لیکن نئے اضافی اصوات کی گنجایش نہ رہی تھی۔ خاص طور پر پنجابی، سرائیکی، پشتو، اور سندھی کے زیر اثر اردو میں جو نئے حروف تہجی سامنے آرہے ہیں یا آسکتے ہیں، ان کے لیے کوئی جگہ نہ تھی۔

اس سے پہلے کہ ہم ضابطہ تختی کے کم سے کم خانے استعمال کرنے کی بات کریں ہم ان اردو صوتیوں پر ایک نظرِ واپسیں ڈال لیں تاکہ ان کی گنجایش مستقبل قریب میں نکالی جاسکے تو بہتر ہوگا۔

سب سے پہلے ہم پنجابی/گورمکھی/لاہوری کی ان آوازوں یا اصوات کو لیتے ہیں جو پاکستانی اُردو بول چال میں اب عام طور پر مستعمل ہیں لیکن لکھنے میں نہیں آپاتے جبکہ اُردو بول چال کی تحقیق میں یہ صوتیے اب اردو کا حصہ بنتے جارہے ہیں۔

ا۔ بھ اور پھ کی درمیانی صوت جو اہل لاہور بھا، بھائی، اور پھول اور بھول کے لیے استعمال کرتے

ہیں۔اس کے لیے ''بھ'' کے اوپر ایک نقطے کی سفارش کی گئی تھی یعنی '' بھ '' اور بعض کے نزدیک ''بھ'' کے نیچے ایک اور عمودی نقطے '' بھ'' کی صورت میں لکھا جانا چاہیے۔ان دونوں کا التباس ''ب'' اور ''ن'' اور بھ'' کے ساتھ ہوسکتا ہے۔ چنانچہ اس کے لیے یہی سفارش ہے کہ ''بھ'' کے نیچے لکیر ''بھ'' کی صورت میں اسے لکھا جائے لیکن پھ کے نیچے لکیر لگا کر نہ لکھا جائے۔کیونکہ اردو میں یہ الفاظ پھ کی نسبت بھ سے بہتر ظاہر ہوتے ہیں۔

گویا ''بھ'' دراصل ''بھ'' کی طرف راجع ہے۔

۲۔ جھ اور چھ کی درمیانی صَوت ''جھا'' کے لیے بھی ''جھ'' کے نیچے ایک افقی لکیر ''جھ'' موزوں ہے۔ یہ آواز بھی چھ سے زیادہ جھ کی طرف راجع ہے۔

۳۔تھ اور دھ کی درمیانی صَوت کے لیے ''دھ'' کے اوپر دو نقطے ''تھ'' اسے ت کی طرح مائل کر کے صوتیہ ظاہر کرنے میں مدد دیں گے، جیسے ''تھوتی''۔ یہ صوت تھ کی نسبت دھ کی طرف زیادہ راجع ہے۔

۴۔ٹھ اور ڈھ کی درمیانی صَوت کے لیے ''ڈھ'' کے اوپر افقی لکیر ''ڈھ'' موزوں ہے جیسے ''ڈھول''۔ یہ صَوت ٹھ کی نسبت ڈھ کی طرف راجع ہے۔

۵۔کھ اور گھ کی درمیانی صَوت ''گھر'' یا ''گھن'' ادائیگی کے لیے ''ق'' پر ایک افقی لکیر ''ق'' کی صورت میں آسانی ادا ہوسکتی ہے۔جیسے ''قر۔قن''۔اس طرح ہم ''کھ۔اور گھ'' کے نیچے دو نقطوں سے بچ سکتے ہیں، جو ان میں ہمارے لیے کتابت، کمپیوٹر کی الجھنیں بڑھاتے ہیں۔

جہاں تک سرائیکی '' ٻ '' کا تعلق ہے وہ بھی ''ی'' کے ساتھ التباس پیدا کرتا ہے، چنانچہ اس کے نچلے نقطے کو افقی لکیر کی صورت میں دیا جائے تو یہ ''ٻ'' کی صورت اختیار کر جاتا ہے۔

سندھی اور ہندکو ''ڑون'' کو ''ڻ'' کے اوپر ط لگا کر ظاہر کیا جاتا ہے۔یہ صَوت سرائیکی میں بھی ملتی ہے۔

تأثر، جرأت، قرأت وغیرہ کے الف ہمزہ کی صَوت ابھی تک معین نہیں تھی۔اسی طرح مؤثر، مؤسسہ میں ''ؤ'' پر ہمزہ کی صوت اور پشتو میں سدوزئی، طوطالئی، اور خانئ وغیرہ جیسی اصوات میں ئ کو بھی الگ ظاہر کرنے کی ضرورت ہے۔یوں أ، ؤ، ئ، کے حروف شامل ہوتے ہیں۔

ان پہلوؤں سے روایتی حروف ۵۶ اور نئے حروف ۲۱ بنتے ہیں۔(دیکھیے نمبر۴: گروہی ترتیب حروف)۔یہ نئے ۲۰ حروف حسب ذیل ہیں۔

أ، ٻ، بھ، جھ، دھ، ڈھ، ٻ، ف، ق، م، نھ، ڻ، ن، ن، ؤ، ۋ، ۆ، ئ، ئ، بھ۔

ابھی اور نہ جانے کتنے حروف پاکستانی اُردو کی اصوات میں شامل ہونے کو بےقرار ہیں۔یاد رہے کہ اس تناظر میں ہم اُردو کو ایک زبان کی حیثیت سے لے رہے ہیں، نہ کہ محض ادبی محضر (Literary

Discourse)۔اس کا بنیادی فریضہ بول چال، ادائیگی و اظہار اور تقریری و تحریری استعمال ہے۔ پاکستانی اُردو اسی سمت میں آگے بڑھ رہی ہے۔

اب اگر کمپیوٹر کی ضابطہ تختی کو دیکھا جائے تو اس کے خانے اس روزافزوں مانگ کا مقابلہ کرنے کی سکت نہیں رکھتے۔ جلدی میں کیا گیا کوئی بھی فیصلہ اُردو کو نہ صرف پابندیوں اور جکڑ بندیوں کا شکار بنا دے گا بلکہ اس کی ترقی اور پھیلاؤ بھی رک کر جمود کی صورت پیدا ہو جائے گی۔ چنانچہ ہمیں یہاں دو بنیادی اصولوں کو ملحوظ خاطر رکھنا ہوگا۔

الف: اردو حروف کی بنیادی شکلوں کی معیار بندی۔

ب: ان شکلوں سے نئے حروف کی تشکیل کے امکانات۔

ان اصولوں کی روشنی میں دیکھیں تو ہمیں اردو میں حروف تہجی کی املا کی اجزا شکنی (Atomizing) کرنا ہوگی اور اپنی ضابطہ تختی کے خانے میں ان اجزاء کو بخشنا ہوں گے۔ سابقہ سیمینار میں یہ بات سامنے آئی کہ اُردو ضابطہ تختی کے ساتھ بہرصورت تمام موجود، نئے اور آئندہ حروف ان اجزاء سے ملا کر وضع ہوتے ہیں، ہو سکتے ہیں یا ہوں گے۔ اب اگر ان کی امتزاجی تکنیک کا گراف ملاحظہ کریں (ضمیمہ ۲)۔ تو ہمیں ان کے ملاپ کی منطقی ترتیب نظر آ جاتی ہے۔ یہاں ہمیں ''ن'' کو '' ں '' سے الگ جز کی حیثیت دینا پڑتی ہے۔ کیونکہ اگر ہم ''ں'' کو ایک بالائی نقطے یعنی ''نمبر ے'' سے ملائیں تو ''ن'' بن جاتا ہے۔ لیکن ڑون کے لیے پھر اسی ''ںٖ'' کو بالائی ط سے ملانے کے لیے ہمیں نمبر ۵ کی طرف واپس آنا پڑتا ہے۔ اور یوں ''ن'' ترتیب Sorting Order میں ''ں'' سے پہلے آ جاتا ہے جو غلط ہے۔ اسی طرح ڻ اور ن، نھ کے بعد آتے ہیں، جو غلط ہیں۔ انھیں بہرحال ن کے بعد اور نھ سے پہلے واقع ہونا چاہیے۔

اس منطقی امتزاج میں ''یھ'' بھی ''ے'' سے پہلے واقع ہوتا ہے اور یوں ''الف تا ے'' کا مقولہ بھی قائم رہتا ہے۔ اس تکنیکی ترتیب میں صرف ایک خرابی ہے کہ ''ح'' ترتیب میں ''ج'' سے پہلے واقع ہو کر ہماری روایتی ترتیب کو ختم کر دیتا ہے۔ اس کا علاج ''ضابطہ تختی'' میں اس طرح ڈھونڈا جا سکتا ہے کہ ہمیں کوئی نہ کوئی محرک (Drive) یا سافٹ ویئر استعمال کرنا پڑے گا۔ چنانچہ دو چشمی ھ کو الگ جز (Atom) مان کر علیحدہ رکھ دیا گیا تھا تا کہ محرک (Engine or Drive) اسے جوڑ کر بیرونی اظہار (External Representation) میں ایک حرف بنا کر پیش کر دے۔ جب حروف کے اجزاء الگ الگ رکھتے ہیں تو ان کے دیگر اجزاء یعنی نقاط اور علامات پر بھی نظر ڈالنی چاہیے۔ اگر ہم حروف تہجی کے خالی املائی اجزاء دیکھیں تو

یہ تعداد میں ۲۲ نظر آتے ہیں۔ جن پر سے نقطے یا علامات ا ب ح د ر س ص ط ع ں ک گ ل م ں و ہ ء ی ھ ے

ان حروف کے آخر میں ھ دراصل ثانوی جز ہے۔ ان کے نقاط اور علامات دیکھیں تو یہ چودہ ہیں۔

ان کے علاوہ زیر، زبر، پیش، شد اور بڑی مد اور دیگر بہت سے اجزاء اور علامات ہیں۔ (دیکھیں ضمیمہ نمبر ا)

اضافی اجزاء کے پہلے خانے یعنی ب، پ، اور الف کے خانے سے پہلے رکھ دیے جائیں، اسی طرح ضابطہ تختی میں ثانوی حروف یا اجزاء ''ء'' ''ھ'' وغیرہ دیگر اضافی علامات وغیرہ کے ساتھ نقاط اور علامات کے بعد رکھ دیئے جائیں (دیکھیے ضمیمہ نمبر ۵)۔

اب ہمارے ۷۶ حروفِ تہجی حسب ترتیب سے واقع ہوتے ہیں۔ قوسین میں نئے مجوزہ حروف ہیں:

ا، (اء)، (آ)، ب (ب) بھ، (بھ)، پ، پھ، ت، ٹھ، ث، ج، جھ، (جھ)، چ، چھ، ح، خ، د، دھ، (دھ) ڈھ، (ڈھ)، ذ، ر، ڑھ، ز، ژ، س، ش، ۔ (س) ص، ض، ط، ظ، ع، غ، ف، ق، (ق) ک، کھ، گ، گھ، ل، لھ، م، (مٖ)، (مھ)، ں، (نھ)، ن، (نٔ)، (ن)، (ن)، وھ، ہ، ۃ، ء، ی، (یٔ)، (یٔ)، (یھ)، ے۔

— قوسین میں دیے گئے حروف کی اصوات سے شاید انکار مشکل ہو۔ اس کی املا پر بھی یقیناً منطقی طور پر اختلاف ممکن نہ ہو سکے گا۔ خاص طور پر تکنیکی حوالوں سے۔ چنانچہ ہمیں ان کی گنجایش ضابطہ تختی میں رکھنا ہو گی۔ خواہ اس کے لیے حروف کو اجزا (Atoms) میں تقسم کرنا پڑے۔ ضابطہ تختی کی عالمی معیار بندی (یونی کوڈ وغیرہ) کی طرف قدم اٹھاتے ہوئے ہمیں ایک بار پھر ان تکنیکی اور مستقبل کی ضروریات کو ملحوظ خاطر رکھنا چاہیے۔ ہمارے تکنیکی ماہرین خاص طور پر اُردو کمپیوٹر سے وابستہ افراد کو اُردو کی ادبی اور اشاعتی ضروریات سے زیادہ اس امر پر توجہ دینی چاہیے کہ اُردو زبان کی بنیادی ضروریات کیا ہیں۔ یعنی What is کو ملحوظ رکھا جائے اور دیکھا جائے کہ اس کے امکانات کا پھیلاؤ (Span) کتنا ہے۔ کل کی ضرورتیں ہمیں آج دیکھ لینی چاہئیں۔ یہ نہ ہو کہ مستقبل قریب میں ہم اپنے جلدی میں کیے گئے فیصلوں پر پچھتائیں۔ یاد رہے کہ اندرونی اظہار (Internal Representation) کچھ اور بات ہے اور ترتیب (Sorting Order) کچھ اور ہے۔ اسی طرح ان کی تلاش (Search) اور ان کی رسائی (Access) کچھ اور شے ہے۔ ہمیں ان میں سے ہر بات کو الگ الگ حوالے سے دیکھنے کی ضرورت ہے۔ کمپیوٹر میں ان کے حوالے سے الگ الگ محرکات اور سافٹ ویئر تیار ہو سکتے ہیں۔

۱-املائی اجزائے حروف

نمبر شمار	خالی اجزاء (۲۲)	ثانوی اجزاء (۲)	+ +اضافی اجزاء (۱۴)
۱-	ا	۲۳+ ـٓ : آ	۲۵ ـ = ب، بھ، ج، جھ
۲	ٮ	۲۴+ ھ = بھ، بھ، پھ	۲۶ ـِ = ب، بھ، جھ
		تھ، ٹھ،	۲۷ ـ۪ = پ، پھ، چ، چھ
۳	ح	+ ھ = جھ، جھ، چھ	۲۸ ـٰ = ت، تھ، ذھ
۴	د	+ ھ = دھ، ڈھ،	۲۹ ـٹـ = ٹ، ٹھ، ڈ، ڈھ، ڑ، ڑھ، ںٹ
		ڈھ، ڈھ	۳۰ ـ = ڈھ، ق
۵	ر	+ ھ = رھ، ڑھ	۳۱ ـٔ = خ، ذ، ز، ض، ظ، غ، ف
۶	س		۳۲ ـ۠ = ث، ژ، ش، ف
۷	ص		۳۳ ـ = بْ
۸	ط		۳۴ ـ = اُ، وُ، یُ
۹	ع		
۱۰	ق		۳۵ ـ = مٚ، نھٚ، وٚ
۱۱	ؤ		۳۶ ـ٘ = ںٛ، وٛ، یٛ
۱۲	ک	+ ھ = کھ	۳۷ ـ = ءٔ
۱۳	گ	+ ھ = گھ	
۱۴	ل	+ ھ = لھ	
۱۵	م	+ ھ = مھ	
۱۶	ں	+ ھ = نھ	
۱۷	ن	+ ھ = نھ	
۱۸	و	+ ھ = وھ	
۱۹	ہ		
۲۰	ء		
۲۱	ی	+ ھ = یھ	
۲۲	ے		

۲-املائی اجزاء کا امتزاج

نمبر	اضافی اجزاء ← / سادہ اجزاء ↓	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶
		ـ.	ـ:	ـ⁘	ـ‥	ط	ـ	ـ.	ـ∴	ـ.ـ	ـر	ـ‥	ـ٧	ـ٨	ـم	ـہ	ھ
۱	ا										ارٛ					آ	
۲	ٮ	ب	ٻ	پ	ت	ٹ			ث								
	ٮھ	بھ	ٻھ	پھ	تھ	ٹھ											ھ
۳	ح	ج		چ				خ									
	حھ	جھ	ڄھ	چھ													ھ
۴	د					ڈ		ذ									
	دھ				ڌھ	ڈھ	ڏھ										ھ
۵	ر					ڑ		ز	ژ								
	رھ					ڑھ											ھ
۶	س								ش	ڛ							
۷	ص							ض									
۸	ط	.						ظ									
۹	ع							غ									
۱۰	ڡ							ف	ڤ								
۱۱	ٯ				ق		ق										
۱۲	ک																
	کھ																ھ
۱۳	گ																
	گھ																ھ
۱۴	ل																
	لھ																ھ
۱۵	م												مٛ				
	مھ																ھ
۱۶	ں																
	ںھ																ھ
۱۷	ن					ڻ								ڻ	نٛ		
	نھ																ھ

۱۸	و										وَ		و٘	و̂			
	وھ																ھ
۱۹	ہ				ۃ												
۲۰	ء																
۲۱	ی										یُ	ئ		ی̂			
	یھ											ئھ					ھ
۲۲	ے																

(نوٹ اس تکنیکی ترتیب میں ح ج سے پہلے آئے گا)۔

۱	ا	۲۷	ڈھ	۵۳	لھ
۲	اُ	۲۸	ذ	۵۴	م
۳	آ	۲۹	ر	۵۵	مٚ
۴	ب	۳۰	رھ	۵۶	مھ
۵	ٻ	۳۱	ڑ	۵۷	ں
۶	بھ	۳۲	ڑھ	۵۸	ںھ
۷	ٻھ	۳۳	ز	۵۹	ن
۸	پ	۳۴	ژ	۶۰	ݨ
۹	پھ	۳۵	س	۶۱	نٛ
۱۰	ت	۳۶	ش	۶۲	نٗ
۱۱	تھ	۳۷	ښ	۶۳	نھ
۱۲	ٹ	۳۸	ص	۶۴	و
۱۳	ٹھ	۳۹	ض	۶۵	ۏ
۱۴	ث	۴۰	ط	۶۶	وٚ
۱۵	ح	۴۱	ظ	۶۷	وٛ
۱۶	ج	۴۲	ع	۶۸	وھ
۱۷	جھ	۴۳	غ	۶۹	ہ
۱۸	ڄھ	۴۴	ف	۷۰	ۃ
۱۹	چ	۴۵	ڤ	۷۱	ء
۲۰	چھ	۴۶	ق	۷۲	ی
۲۱	خ	۴۷	ڨ	۷۳	یٚ
۲۲	د	۴۸	ک	۷۴	یٛ
۲۳	دھ	۴۹	کھ	۷۵	یھ
۲۴	ڏھ	۵۰	گ	۷۶	ے
۲۵	ڈ	۵۱	گھ		
۲۶	ڈھ	۵۲	ل		

۴-گروہی ترتیبِ حروف

	روایتی حروف (۵۶)	نئے حروف (۲۰)
۱	ا	۲-اُ
۳	آ/اٗ	
۴	ب	۵-بٖ
۶	بھ	۷-بھٖ
۸	پ	
۹	پھ	•
۱۰	ت	
۱۱	تھ	
۱۲	ٹ	
۱۳	ٹھ	
۱۴	ث	
۱۵	ج	
۱۶	جھ	۱۷-جھٖ
۱۸	چ	
۱۹	چھ	
۲۰	ح	
۲۱	خ	
۲۲	د	
۲۳	دھ	۲۴-دٖھ
۲۵	ڈ	
۲۶	ڈھ	۲۷-ڈٖھ
۲۸	ذ	
۲۹	ر	
۳۰	رھ	•
۳۱	ڑ	
۳۲	ڑھ	
۳۳	ز	
۳۴	ژ	
۳۵	س	
۳۶	ش	۳۷-یش
۳۸	ص	

	روایتی حروف	نئے حروف
۳۹	ض	
۴۰	ط	
۴۱	ظ	
۴۲	ع	
۴۳	غ	
۴۴	ف	۴۵-ڤ
۴۶	ق	۴۷-قٖ
۴۸	ک	
۴۹	کھ	
۵۰	گ	
۵۱	گھ	
۵۲	ل	
۵۳	لھ	
۵۴	م	۵۵-مٚ
۵۶	مھ	
۵۷	ن	۵۸-نھٚ
۵۹	ں	۶۰-ںٛ
		۶۱-ںٛ
		۶۲-ںٛ
۶۳	نھ	
۶۴	و	۶۵-وَ
		۶۶-وٚ
		۶۷-وٛ
۶۸	وھ	
۶۹	ہ	
۷۰	ۃ	
۷۱	ء	
۷۲	ی	۷۳-یٗ
		۷۴-یٛ
		۷۵-یھ
۷۶	ے	

۳-تکنیکی ترتیبِ حروف

۵- ضابطہ تختی میں ترتیبِ اجزاء

۱	ـــ ہ	۱۱	ط	۲۱	و	۳۱	ـبـ
۲	ـــ خ	۱۲	ع	۲۲	ہ	۳۲	ـــر
۳	ـــ پ	۱۳	ف	۲۳	ء	۳۳	ـتـ
۴	ا	۱۴	ق	۲۴	ی	۳۴	ـںـ
۵	ب	۱۵	ک	۲۵	ے	۳۵	ــ^
۶	ح	۱۶	گ	۲۶	ـثـ	۳۶	ـحـ
۷	د	۱۷	ل	۲۷	ـٹ	۳۷	ـ~ـ
۸	ر	۱۸	م	۲۸	ـــ	۳۸	ھ
۹	س	۱۹	ں	۲۹	ـنـ	۳۹	ـسـ
۱۰	ص	۲۰	ن	۳۰	ـشـ	۴۰	ــ

("خبارِ اُردو"، اسلام آباد، دسمبر ۱۹۹۹ء)

۴

اُردو میں انٹرنیٹ وغیرہ کے لیے قابلِ رُجوع نقلِ حرفی

اُردو میں نقل حرفی (Transliteration)، خاص طور پر رومن حروف میں لکھنے، کے لیے اب تک بیس سے زائد نظام وجود میں آ چکے ہیں۔ ان میں برٹش لائبریری، لائبریری آف کانگریس، انسائیکلوپیڈیا آف اسلام کے نظام خاص طور پر معروف ہیں۔ لیکن ان میں سے ایک نظام بھی ایسا نہیں جسے رجوع یا واپسی کے قابل (Reversible) سمجھا جا سکے۔ خاص طور پر کمپیوٹر پر استعمال کے لیے کسی ایسے نظام کی ضرورت محسوس ہوتی ہے۔ جو رومن حروف سے اُردو میں نقل حرفی کرتے ہوئے صحیح ترین صورت اختیار کرے اور اس میں رومن حروف کے اوپر یا نیچے علامات بھی نہ لگانا پڑیں۔

کمپیوٹر میں اُردو کے استعمال کے لیے بعض امور کو طے کرنے کے سلسلے میں ایسے بیسیوں تکنیکی مسائل کا سامنا کرنا پڑتا ہے۔ اس سے پہلے اُردو کے ماہرین اور اہل علم کو کبھی ان سے واسطہ نہ پڑا اور نہ ان کے لیے کوئی حل سوچا گیا۔

ظاہر ہے کہ انسانی علمی تقاضے کچھ اور ہوتے ہیں اور خالص تکنیکی مسائل کسی اور سمت اشارہ کرتے ہیں۔ اکیسویں صدی کے آغاز سے پہلے کسی نے یہ سوچا بھی نہ تھا کہ زبانیں سماجی میل جول اور موضوعی (Subjective) رویوں سے ہٹ کر تکنیکی اور معروضی (Objective) حالات کی بھینٹ چڑھ جائیں گی اور تہذیبی سطح پر صرف وہی زبان قائم رہے گی جو ٹیکنالوجی کے ہاتھوں ذبح ہونے سے بچ رہے گی۔

آج بھی بہت سے لوگ یہی سمجھتے ہیں کہ زبان محض بول چال کی چیز ہے، شعر و ادب کی ملکیت ہے، وہ بھلا کب فنا ہو سکتی ہے، خاص طور پر کمپیوٹر کے ہاتھوں جبکہ کئی زبانوں کے لفظ کار word processor تک وجود میں آ چکے ہیں۔ پھر بھلا زبانوں کو کیا خطرہ ہے۔ حقیقت یہ ہے کہ کمپیوٹر کاری (Computing) کے لیے خاص طور پر کوائفی اساس (Database) اور انٹرنیٹ کی دنیا میں تمام زبانوں کو آج رومن حروف کے ساتھ ایک عجیب سی تکنیکی ہم آہنگی پیدا کرنے کی ضرورت درپیش ہے جو عام سماجی صورت حال میں کسی کے حاشیۂ خیال میں بھی نہیں آ سکتی۔

وجہ صاف ظاہر ہے کہ کمپیوٹر انگریزی دانوں کی ایجاد ہے۔ انھوں نے اس کی بنیاد میں رومن حروف

رکھ دیے ہیں۔ ہر زبان کو اپنے حروف/اصوات/علامات انھی حروف اور علامات کے متبادل کے ساتھ پیش کرنے کی ضرورت ہے۔اس ضمن میں متعدد ایسے تکنیکی مسائل سامنے آئے ہیں۔ جب ان کے حل کے لیے اُردو کے ذخیروں کو کھنگالا جاتا ہے تو وہاں ایک طویل اور گہری خاموشی ملتی ہے،اس سب کے جواب میں۔

موضوعیت کوئی برخود غلط شئے نہیں۔ ذہن (Mind) کی داخلیت اسی کے باعث تخلیق کے مراحل طے کرتی ہے۔رضا اور مرضی اسی کے نتیجے میں وجود میں آتے ہیں۔لیکن کمپیوٹر نہ تو تخلیقی کام انجام دیتا ہے۔اور نہ من مرضی کے کام کرتا ہے۔اس کے برعکس انسان ایک ذہین مخلوق ہے۔ یوں کہیے کہ وہ اپنی مرضی کا مالک ہوتا ہے۔اس میں موضوعیت ہوتی ہے۔ ذہانت ہوتی ہے۔اس کا سب سے بڑا ثبوت یہ ہے کہ انسان دھوکا دیتا ہے اور کھاتا ہے، غلطی کرتا اور اس کی تصیح کرتا ہے، خطا کرتا ہے اور اس کا ازالہ کرتا ہے، مگر کمپیوٹر کو آپ خواہ ''مصنوعی ذہین'' قرار دیں لیکن وہ یہ سب کام نہیں کر سکتا۔مثل مشہور ہے کہ کمپیوٹر میں کوڑا ڈالو،کوڑا ہی نکلے گا یعنی اس سے وہی کچھ برآمد ہوگا جو اس میں داخل کیا جائے گا۔

اس حوالے سے دیکھیں تو اُردو حروف کو رومن صورت میں ڈھال کر جب کمپیوٹر میں بھیجا جائے گا تو سکرین پر واپسی یا رجوعی عمل میں، جب کہ ہمیں اُردو حروف نظر آنے کی توقع ہو، کمپیوٹر کو وہی رومن اُردو میں ڈھال کر پیش کرنے چاہئیں۔لیکن ابھی تک نقلِ حرفی کا ایسا کوئی نظام سامنے نہیں آسکا جو اس واپسی یا رجوعی عمل کو درست رکھ سکے یا موضوعیت کی بجائے معروضیت پر مبنی ہو۔

اس عمل کو یوں بھی کہہ سکتے ہیں کہ کمپیوٹر ہر نوع کے لیے الگ حرف یا علامت طلب کرتا ہے لیکن اُردو حروف کو رومن بنانے کے لیے ماہرین نے کوئی تکنیکی طریقہ ایجاد نہیں کیا تھا۔مثلاً اگر ہم ''کعبہ'' کو رومن میں لکھیں تو "Ka'ba","Kaba", "Kaaba" کی سی صورتیں اختیار کی جاتی ہیں۔جن میں a کا حرف کبھی الف، کبھی زبر، کبھی عین اور کبھی ''ہائے مختفی'' کے لیے استعمال ہوتا ہے۔اب کمپیوٹر تو اتنا ذہین نہیں کہ انھیں دوبارہ اُردو میں لکھتے ہوئے آپ کا مطلوبہ اُردو لفظ ''کعبہ'' پیش کر دے۔ وہ لازماً لفظ کو الٹ پلٹ کر دے گا یا پھر ''پریشان اور پاگل'' ہو جائیگا۔ یہی صورت دوسرے صوتی طور پر ہم آہنگ حروف کی ہے مثلاً ز، ذ، ض، ظ کے لیے ایک ہی رومن حرف z استعمال کیا جاتا ہے۔نقل حرفی کے بعض طریقوں میں Z کے اوپر نیچے نقطے استعمال کیے جاتے ہیں۔''ض'' کے لیے "dh" ''دھ'' کے لیے بھی استعمال ہوتا ہے اور کہیں ''ذ'' کے لیے مستعمل ہے۔انٹرنیٹ میں رومن حروف کے اوپر نیچے علامات استعمال نہیں ہوسکتیں۔ یہی صورت مشدد حروف کی ہے۔انھیں رومن حروف میں دوبارہ لکھ دیا جاتا ہے۔مثلاً ''بّ'' کے لیے bb۔اسے اُردو میں واپس تبدیل کرتے ہوئے کمپیوٹر ''بب'' بنا دے گا۔ چنانچہ جب اُردو، عربی، فارسی کے حروف کی نقل حرفی کی موجودہ سکیموں کا جائزہ لیا گیا تو مندرجہ ذیل اصولوں کی روشنی میں وہ آئندہ تکنیکی ترقی کا ساتھ دینے سے معذور ثابت ہوئیں۔

۱۔ اُردو کے ہر حرف اور اس کی صورت یا حرکت کے لیے الگ حرف یا علامت موجود ہو اور وہ کسی دوسرے حروف سے اشتباہ پیدا نہ کرے۔

۲۔ تمام متبادل رومن حروف یا علامتیں قابلِ رجوع یا واپسی ہوں یعنی رومن سے اُردو میں منتقل کرتے ہوئے غلطی کا شائبہ بھی نہ ہو۔

۳۔ تمام حرکات علت (Vowel) کو بھی حروف تسلیم کرتے ہوئے ان کے لیے الگ رومن متبادل حروف یا علت تلاش کی جائے۔

۴۔ رومن متبادل حروف کے لیے صرف انگریزی کے ۲۶ حروف استعمال کیے جائیں اور ان پر امتیاز کے لیے کوئی نشان، علامت یا لکیر نہ دی جائے۔

۵۔ بعض ہم صَوت یا قریب الصوت اُردو حروف کی مختلف صورتیں کے لیے ایک ہی انگریزی حرف کے استعمال میں امتیاز کے لیے حرف کے دائیں طرف کوئی امتیازی علامت استعمال کی جائے۔ جو مفرد صَوت میں کسی اور حرف یا صَوت کے لیے استعمال نہ ہوئی ہو اور موجودہ انگریزی کلیدی تختے پر موجود ہو نیز انٹرنیٹ اور ویب سائٹ کے کنٹرول / پروگرام وغیرہ میں بہت کم مستعمل ہو۔

ان اصولوں پر اُردو حروف کی رومن نقل حرفی کا جائزہ لیا گیا تو معلوم ہوا کہ الف، و، ہ، ء، ع، ی وغیرہ حروف علت (Vowel) کے طور پر بھی استعمال ہوتے ہیں اور ان کی کم وبیش صورتیں اعراب وحرکات کی صورت اختیار کر جاتی ہے۔ ان سب میں امتیاز ملحوظ رکھا جائے تو الف کی دس قسمیں، انواع یا صورتیں سامنے آتی ہیں۔ ''واؤ'' کی ۱۳، ''ہ'' کی ۷، اور ''ء'' کی ۴، ''ع'' کی ۵، ''ی'' ۱۴ صورتیں ہیں۔ ان سب کے تعین کے لیے الگ الگ رومن حروف یا علامتیں درکار ہیں۔

باقی اُردو حروف میں سے ''ف'' اور ''ق'' کی ۳، ۳ صورتیں ہیں۔ پ، ج، چ، ل م کی ۴، ۴ صورتیں ''ب'' کی پانچ، ک اور گ کی ۶، ۶، ''ن'' کی ۷، ''د''، ر، ز'' ''س'' کی ۸، ۸ اور ت کی ۱۰ صورتیں ہیں۔ ان کے علاوہ دیگر حروف انھی گروہوں میں پناہ لیے ہوئے ہیں۔ ان سب کو الگ الگ مشخص کرنا تا کہ ایک بے جان معروضی مشین ان میں امتیاز کر سکے لازم ہے۔

اگر ذرا سا غور کیا جائے تو ''الف'' کی ابتدائی صورت ہے جو حرف صحیح (Consonent) کے طور پر استعمال ہوتی ہے اور ایک آخری صورت ہے، جو لفظ، ہجا یا جوڑ میں بطور علت استعمال ہوتی ہے۔ رومن علامت کی صورت میں ان دونوں میں امتیاز ضروری ہے۔ ''زبر'' دراصل نصف الف علت ہے۔ کھڑی زبر اس سے ذرا بڑی علامتِ علت ہے۔ پھر جب الف زبر، الف زیر اور الف پیش استعمال میں آئیں تو دراصل یہ زبر، زیر پیش ہی ہوتے ہیں، جو لفظ کے شروع میں آتے ہیں لیکن کسی حرفِ صحیح کے بعد آنے سے ذرا

مختلف ہوتے ہیں۔ الف کی ایک صورت مخلوط ہے جو ''بالکل'' جیسے لفظ میں اُردو، عربی میں تو لکھنے میں آتی ہیں لیکن اسے رومن کرتے ہوئے غائب ہو جاتی ہے۔ یعنی "Bilkul"۔ اگر کمپیوٹر اسے دوبارہ اُردو یا عربی میں لکھنا چاہے تو ایسا کرنے سے معذور ہو جاتا ہے۔ یہی صورت حرف ''آ'' کی ہے اگر اسے دو الف مان کر ''ā'' لکھا جائے تو واپسی میں ''اا'' ملیں گے۔

واو کی صورت تو اس سے بڑھ کر ہے۔ واو زبر، زیر، واو پیش میں تو کوئی مشکل نہیں کیونکہ یہ حرکات الف کے برعکس واو صحیح کیساتھ ہیں۔ مگر خود واو علت کی صورت میں اس کی آواز ''O'' کی ہو جاتی ہے۔ زبر کے بعد واو یعنی واو لین کی صورت میں ''ao'' کی ہو جاتی ہے۔ جبکہ ہم ''a'' کو الف صحیح قرار دے چکے ہیں۔ پیش کی علامت دراصل نصف o ہے۔ الٹی پیش اس سے ذرا بڑھ کر ہے۔ پیش واو ایک اور طرح کا واو لین ہے۔ اس کے علاوہ ایک دہرے لین یا تہری علت کی صورت بھی واو کیساتھ در پیش ہے۔ اور وہ ''ہو'' ''جو'' ''خواجہ'' جیسے الفاظ میں ''پیش، واو، الف'' کی ہے۔ واو مخلوط یا واو معدولہ بھی دو ایسی صورتیں ہیں جو فی الوقت اُردو میں ایک ہی صورت ''معدولہ'' کے نام سے معروف ہیں۔ مشدد واو بھی اپنا امتیاز طلب کرتا ہے جو واو صحیح ہی کو دہرانے کا نام ہے۔

Wh کا ایک متبادل ''وھ'' بھی وھیل اور دھائیٹ کی صورت میں موجود ہے۔ مسئلہ اس کے وجود کے اقرار یا انکار کا نہیں، مجموعی کارپس (Corpus) صورت میں کمپیوٹر ٹیکنالوجی میں تمام امکانی صورتیں پیش نظر رکھنا ہوتی ہیں کہ اگر ایسا لکھنا پڑ جائے تو کیا ہو۔

''ۃ'' کا مسئلہ اور بھی گھمبیر ہے۔ اس کی دوسری صورت علت (Vowel) کی ہے جو ''کعبۃ'' راجۃ'' وغیرہ کے آخر میں ''الف'' کی سی آواز دیتی ہے۔ ''ۃ'' مشدد میں پہلے ہائے علت (ہائے مختفی) اور پھر ہائے صحیح کی صورت ابھرتی ہے۔ ''ح'' کی صوت اور شکل کا امتیاز بھی ضروری ہے۔ اور ''ح'' مشدد کی صورت میں بھی علیحدہ متبادل درکار ہے۔ گول ت یعنی ''ۃ'' کی اُردو میں ایک الگ صورت ہے۔ یہ لکھی جاتی ہے لیکن بولی نہیں جاتی یعنی عربی ''ۃ'' میں گول ت کی بجائے صرف ''ت'' کی ذرا سی بھاری صوت ہے۔ دو چشمی ھ کو اُردو میں ''ہ'' کی قائم مقامی نہیں بلکہ بھاری اصوات ''بھ'' پھ، تھ، کے تحریری جز و نصف کی حیثیت حاصل ہے۔

ان سب کے لیے امتیازی رومن حروف یا علامات درکار ہیں نیز کمپیوٹر کو یہ بتانا بھی مقصود ہے کہ اُردو میں ھ الگ طور سے کہیں استعمال نہیں ہوتا، اگر ایسا ہو تو وہ ''ہ'' ہے۔

ہمزہ بنیادی طور پر عربی کا ''الف'' ہے لیکن اُردو میں ''کوئی'' ''کئی'' وغیرہ میں الگ الگ ''حرف'' کے طور پر آتا ہے۔ بعض افراد ہمزہ اضافت کو بھی ''ہمزہ'' سمجھتے ہیں۔ اس پر بحث آگے آتی ہے۔ عین کا حرف ''ہمزہ'' اور الف سے مشابہ ہے۔ خاص طور پر مشدد صورت میں دو الف کی حثیت اختیار کر جاتا

ہے۔

یائے صحیح اور یائے علت کے علاوہ بھی اس کی کئی صورتیں امتیازی حثیت رکھتی ہیں۔ مثلا درمیانی ''ی'' مشدد سے مختلف ہے جو بیک وقت حرف صحیح بھی ہے اور حرفِ علت بھی۔ زیر اور کھڑی زیر بھی نصف اور نیم ''ی'' کی صورتیں ہیں۔ بڑی ''ے'' ایک الگ حیثیت رکھتی ہے اور اس کی لین صورت ''زبر+ ے'' کے طور پر جدا مستعمل ہے۔ ''ژ'' بھی دراصل ''یھ'' کی ایک صورت ہے۔

یہ تمام حروف آپس میں عمل کرتے ہیں تو ان کے لیے بھی بعض قاعدے وجود میں آتے ہیں۔ تکنیکی لحاظ سے ا، د، ء، ع، ی میں باہمی فرق لازم ہے۔ پھر ایسی علامات استعمال کی جائیں جو ان میں علت کے لیے یکسانیت پیدا کریں۔ اسی طرح کھڑی زبر، کھڑی زیر، اور کھڑی پیش کے لیے قاعدہ درکار ہے۔

اُردو کا مسئلہ یہ ہے کہ اس میں حروف انگریزی کی طرح محض پاس پاس رکھ دینے سے نہ تو لفظ تشکیل پاتا ہے اور نہ ایک ترسیمہ Ligature میں آجانے سے معلوم ہوتا ہے۔ کہ وہاں ا، ء، ہ، و، ی، ے بطور حرف صحیح استعمال ہوئے ہیں۔ یا بطور حرف علت۔ مثلاً ''کور'' اور ''کَور'' میں کو کے پہلے ترسیمے واؤ علت ہے اور دوسرے ترسیمے میں واو صحیح۔ چنانچہ رومن نقل حرفی کرنے والے سافٹ ویئر کو مطلع کرنا پڑے گا۔ کہ ''واؤ'' کی کلید برائے ''علت'' رکھی جائے جو حرف سے پہلے یا بعد میں استعمال کی جائے۔ یہی صورت حرف مخلوط کی ہے اس کے لیے بھی ایک الگ کلید مزید استعمال کرنا ہوگی۔

دیگر حروف میں بعض ایسی صورتیں سامنے آتی ہیں جن کے لیے عام یا بنیادی طور پر ایک ہی رومن حرف استعمال ہوتا ہے۔ مثلات، ٹ، ٹھ، ط، و، وغیرہ یا پھر د، دھ، ڈھ، اسی طرح ر، رھ، ڑھ وغیرہ نیز ز، ذ، ض، ظ کی مثال پہلے ہی دی جا چکی ہے۔ س، ث، ص، ش کا مسئلہ بھی کچھ ایسا ہی ہے، ک، کھ اور خ میں مشابہت موجود ہے جیسی کہ گ، گھ اور غ میں ہے۔ ن اور ں کا امتیاز بھی لازم ہے۔ بہت تفحص اور تفکر کے بعد مندرجہ ذیل اصول ہاتھ آئے:۔

ا۔ مجوزہ رومن حروف یا علامتیں پہلے ہی کثرت سے رومن اُردو یا نقل حرفی میں مستعمل ہوں تاکہ اجنبیت یا نئے پن کا احساس کم سے کم ہو۔

۲۔ حروف کی انواع و اقسام یا مختلف صورتوں میں امتیاز کے لیے جو اضافی علامتیں استعمال ہوں وہ قابل خواندگی ہوں یعنی ایسے رومن لفظ کو پڑھنے میں انسانی آنکھ دقت محسوس نہ کرے اور معمولی سی کوشش سے اسے پڑھا جا سکے۔

۳۔ مجوزہ رومن متبادلات سابقہ تمام اصولوں/نظاموں کی جگہ لے کر ایک جامع، عمومی اور معیاری نظام کے طور پر سامنے آئیں۔ نیا نظام تجویز کرتے ہوئے اُردو کے ستر اسّی حروف اور حرکات پیش کرنے کے لیے انگریزی کے ۲۶ حروف تو ناکافی ہیں۔ خاص طور پر x, v, c وغیرہ کو نکال کر ان کی تعداد ۲۳ رہ جاتی ہے۔

ایک تجویز یہ بھی تھی کہ انٹرنیٹ کے لیے حرف کی مختلف صورتوں کا تعین کرتے ہوئے نمبر دے دیئے جائیں مثلاً ذ، ز، ظ، ض اور ان کی مشدد صورتوں کے لیے نمبر دیے جائیں۔ ز=z1 ,z2= زّ , z3= ذ z4 , ذّ =, ض z5 , z6= ضّ z7 =, ظ =z8= ظّ۔

اس طریقے میں دو قباحتیں تھیں۔

۱۔ اُردو حروف کی بعض صورتیں ۱۳ تک پہنچتی ہیں اور اعداد ۹ تک ہیں۔

۲۔ اس طرح سے لکھا ہوا رومن لفظ عام فرد کے لیے قابل خواندگی نہیں۔ اگر چہ اس رومن متبادل کو صرف تکنیکی ماہرین ہی استعمال کریں گے لیکن اگر عام افراد ان رومن حروف کو سکرین پر دیکھنا چاہیں تو کیا ہو؟

۳۔ اگر عبارت کے اندر اعداد ہوں اور انھیں پیش کرنا ضروری ہو تو حروف اور اعداد میں امتیاز کی کیا صورت ہو گی۔

۴۔ انگریزی کے ۲۳ حروف کے ساتھ کون کون سی علامتیں استعمال کی جائیں۔ جو کلیدی تختے پر بھی موجود ہوں لیکن انٹرنیٹ کی پروگرامنگ میں بہت کم مستعمل ہوں۔

بہت غور وخوض کے بعد انگریزی سے ۲۴ حروف (جن میں c بھی مرکب صورت میں چ کے لیے Ch کے طور پر استعمال کیا جاتا ہے۔) اور پانچ علامات !،\،*،^،~ وغیرہ ہاتھ لگیں۔ ''ا، و، ی'' کی مخلوط صورت یا پھر چ، خ، ش، اور غ کی مخلوط صورت بھ، پھ، تھ کی مخلوط صورت سے مختلف ہے۔ اس لیے مخلوط صورتوں بھ، پھ کے لیے ! کی علامت کا اضافہ کیا گیا ہے۔ ا، و، ی کی مخلوط صورت تو کم ہی سامنے آتی ہے۔ مثلاً ''بالکل'' ''خورشید'' اور ''کیا''، کیوں، پیار وغیرہ میں۔ "^" کی علامت کو آگے کھینچی جانے والی اصوات کے لیے اور "~" کی علامت کو اوپر یا گول آواز کھینچنے کی اصوات کے لیے استعمال کیا گیا۔ "*" کو شد کی علامت کا متبادل بنایا گیا تاکہ حروف کو دہرانے کے عمل سے بچا جا سکے۔ یوں اُردو کے صوتی اور مکتوبی ہر دو انداز ابجد کو ملحوظ رکھا گیا نیز مستقبل کے امکانی حروف کے لیے بھی گنجایش پیدا ہو گی۔

اضافی علامات کے استعمال میں باہمی یکسانیت بھی ملحوظ رکھی گئی مثلاً '':'' کی علامت کھڑی زبر، کھڑی زیر، اور الٹی پیش کے لیے بھی استعمال ہوئی۔ ''~'' کی علامت ''مد'' کے علاوہ ٹ، ڈ، ڑ، ن کی ط کی علامت اور ''ہ'' اور ''ح'' اور ی'' اور ''ے'' ''ز'' اور ''ظ'' میں امتیاز کے لیے استعمال ہوئی ہیں۔ چ، خ، ش، غ کے لیے علی الترتیب Gh, Sh, Kh, Ch استعمال ہوتے ہیں لیکن کمپیوٹر انھیں ''ہC'' ''ک ہ'' ''س ہ'' اور ''گ ہ'' پڑھ سکتا ہے۔ چھ، کھ، سھ، گھ بھی اسی طرح لکھے جا سکتے ہیں۔ اگر ان کے ساتھ کوئی امتیازی نشان مہیا کر دیا جائے جو ''`'' کی صورت میں دے دیا گیا ہے تو سافٹ ویئر اسے چ، خ، ش اور غ کی حیثیت سے شناخت کر لے گا۔

اضافت ایک ایسا مسئلہ ہے جو کبھی ء اور کبھی زیر کی صورت اختیار کر کے ہمزہ اور زیر کی ساتھ مغالطہ پیدا کرتا ہے۔ حالانکہ اشاریاتی ترتیب (Collating Order) میں ہمزہ اور زیر کو تو مقام حاصل ہے مگر

Version 0.5, October 17, 2000

Reversible Urdu <--->Transliteration
to be used in
Computers/ E-mail/ Internet

قابلِ رُجوع اُردُو نقلِ حرفی
برائے
کمپیوٹر / انٹرنیٹ

(الف) صوتی حرفی گروہ اور متبادلات

گروہ نمبر	حرف	رومن متبادل	کیفیت
1	ا	a	الف ابتدائی
	ـا	a^	آخری (Vowel)
	ءا	'a!	ہمزہ الف
	أ	a'!	الف ہمزہ
	ـَ	e	زبر
	ـٰ	a`	کھڑی زبر
	اَ	a/ae	الف زبر
	اِ	i/ai	الف زیر
	اُ	u/au	الف پیش
	↑	a!	الف مخلوط
	آ	a~	
	ـً	ee	تنوین
2	و	w	ابتدائی
	وَ	we	
	وِ	wi	
	وُ	wu	
	ـــو	o	واو مجہول (Vowel)
	ـَو	eo	لین
	ـُ	u	پیش
	ـٗ	u`	اُلٹی پیش
	ـُو	uo	واو معروف
	ـُوا	uoa^	دہرا لین
	ؤ/و	u!	مخلوط / معدولہ
	وّ	w*	مشدد
	وھ	wh!	
	وٌ	uu`	تنوین
3	ہ	h	
	ہ	h^	ہائے مختفی (Vowel)
	ہّ	h*	
	ح	h~	
	حّ	h~*	
	ۃ	ht!	
	ھ	h!	
4	ء	'	
	ءَ	'e	
	ءِ	'i	
	ءُ	'u	
5	ع	'^	
	عَ	'^e	
	عِ	'^i	
	عُ	'^u	
	عّ	'^*	
6	ی / یـ	y	ابتدائی
	یَ	ye	
	یِ	yi	
	یُ	yu	
	یـ / ی	iy	درمیانی آخری (Vowel)
	یّ	iy*	مشدد
	ی	y!	مخلوط
	ـِ	i	زیر
	ـٖ	i`	کھڑی زیر
	ـٍ	ii	تنوین
	یھ	yh!	
	ے	y~	
	ـَے	ey~	لین
	ئ	y^	
7	ب	b	
	بّ	b*	
	بھ	bh!	
	بھّ	bh!*	
8	پ	p	
	پّ	p*	
	پھ	ph!	
	پھّ	ph!*	

	t	ت	9
	t*	تّ	
	th!	تھ	
	th!*	تّھ	
	t~	ٹ	
	t~*	ٹّ	
	t~h!	ٹھ	
	t~h!*	ٹّھ	
	t^	ط	
	t^*	طّ	
	j	ج	10
	j*	جّ	
	jh!	جھ	
	jh!*	جّھ	
	ch`	چ	11
	ch`*	چّ	
	ch`h!	چھ	
	ch`h!*	چّھ	
	d	د	12
	d*	دّ	
	dh!	دھ	
	dh!*	دّھ	
	d~	ڈ	
	d~*	ڈّ	
	d~h!	ڈھ	
	d~h!*	ڈّھ	
	r	ر	13
	r*	رّ	
	rh!	رھ	
	rh!*	رّھ	
	r~	ڑ	
	r~*	ڑّ	
	r~h!	ڑھ	
	r~h!*	ڑّھ	
	z	ز	14
	z*	زّ	
	z+	ذ	
	z+*	ذّ	
	z^	ض	
	z^*	ضّ	
	z~	ظ	
	z~*	ظّ	
	s	س	15
	s*	سّ	
	s~	ث	
	s~*	ثّ	
	s^	ص	
	s^*	صّ	
	sh`	ش	
	sh`*	شّ	
	f	ف	16
	f*	فّ	
	q	ق	17
	q*	قّ	
	k	ک	18
	k*	کّ	
	kh!	کھ	
	kh!*	کّھ	
	kh`	خ	
	kh`*	خّ	
	g	گ	19
	g*	گّ	
	gh!	گھ	
	gh!*	گّھ	
	gh`	غ	
	gh`*	غّ	
	l	ل	20
	l*	لّ	
	lh!	لھ	
	lh!*	لّھ	

21	م مّ مھ مھّ	m m* mh! mh!*	
22	ن نّ نھ نھّ ں ںھ نؕ	n n* nh! nh!* n^ n^h! n~	
23	اضافت ِ	-i-	
24	تشدید ّ	*	
25	مدّ ٓ	!ٓ	مدّ
26	لا	x	
27	ۂ	c	
28	لفظ مکمل	Space Bar	
29	اردو شروع	[x x]	

(ب) اردو حروف تہجی کی ترتیب

1- ا (الف ابتدائی)	a
2- ا (الف آخری۔ Vowel)	a^
3- ء (ہمزہ الف)	'a!
4 ۔أ (الف ہمزہ)	a'!
5- ـَ ۔(زبر)	e
6- ـٰ ۔(کھڑی زبر)	a`
7- اَ (الف زبر)	a/ ae
8- اِ (الف زیر)	i/ai
9- اُ (الف پیش)	u/au
10- آ (الف مخلوط)	a!
11- آ	a~
12- تنوین ً ٍ ٌ	e*
13- ب	b
14- بّ	b*
15- بھ	bh!
16- بھّ	bh!*
17- پ	p
18- پّ	p*
19- پھ	ph!
20- پھّ	ph!*
21- ت	t
22- تّ	t*
23- تھ	th!
24- تھّ	th!*
25- ٹ	t~
26- ٹّ	t~*
27- ٹھ	th!
28- ٹھّ	th!*
29- ث	s~
30- ثّ	s~*
31- ج	j
32- جّ	j*
33- جھ	jh!
34- جھّ	jh!*
35- چ	ch`
36- چّ	ch`*
37- چھ	ch`h!
38- چھّ	ch`h!*

نمبر	حرف	کوڈ
39	ح	h~
40	حّ	h~*
41	خ	kh`
42	خّ	kh`*
43	د	d
44	دّ	d*
45	دھ	dh!
46	دّھ	dh!*
47	ڈ	d~
48	ڈّ	d~*
49	ڈھ	d~h!
50	ڈّھ	d~h!*
51	ذ	z!
52	ذّ	z!*
53	ر	r
54	رّ	r*
55	رھ	rh!
56	رّھ	rh!*
57	ڑ	r~
58	ڑّ	r~*
59	ڑھ	r~h!
60	ڑّھ	r~h!*
61	ز	z
62	زّ	z*
63	ژ	y^
64	ژّ	y^*
65	س	s
66	سّ	s*
67	ش	sh`
68	شّ	sh`*
69	ص	s^
70	صّ	s^*
71	ض	z^
72	ضّ	z^*
73	ط	t^
74	طّ	t^*
75	ظ	z~
76	ظّ	z~*
77	ع	'^
78	عَ	'^e
79	عِ	'^i
80	عُ	'^u
81	عّ	'^*
82	غ	gh`
83	غّ	gh`*
84	ف	f
85	فّ	f*
86	ق	q
87	قّ	q*
88	ک	k
89	کّ	k*
90	کھ	kh!
91	کّھ	kh!*
92	گ	g
93	گّ	g*
94	گھ	gh!

95- گھؔ	gh!*
96- ل	l
97- لؔ	l*
98- لھ	lh!
99- لھؔ	lh!*
100- م	m
101- مؔ	m*
102- مھ	mh!
103- مھؔ	mh!*
104- ن	n
105- نؔ	n*
106- نھ	nh
107- نھؔ	nh!*
108- ں	n^
109- ںھ	n^h!
110- نٔ	n~
111- و (ابتدائی)	w
112- وَ	we
113- وِ	wi
114- وُ	wu
115- ـــو (مجہول۔Vowel)	o
116 ۓ و (لین)	eo
117- ـُـ (پیش)	u
118- ـٗـ (الٹی پیش)	u`
119- ـُ و (معروف)	uo
120-ُوا (دہرالین)	uoa^
121- ؤ/وٕ (مخلوط/معدولہ)	u!
122- وؔ	w*
123- وھ	wh!
124- ـٌ تنوین	uu!
125- ہ	h
126-ہ(ہائے مختفی Vowel)	h^
127- ہؔ	h*
128- ۃ	ht!
129- ء	'
130- ءَ	'/ 'e
131- ءِ	'i
132- ءُ	'u
133- ی/یہ(ابتدائی)	y
134- یَ	ye
135- یِ	yi
136- یُ	yu
137- ـیـ/ی(درمیانی/آخری۔Vowel)	iy
138- یؔ	iy*
139- ئ (مخلوط)	y`
140- ـِـ (زیر)	i
141- ـٖـ (کھڑی زیر)	i!
142- ـٍـ تنوین	ii
143- یھ	yh!
144- ے	y~
145- ۓ ے(لین)	ey~
146- ھ	h!
147- ـٓ	*
148- ـٓمد	~!
149- ـِ/ء اضافت	-i-
150- لا	x

151-	ہ	C
152-	لفظ مکمل	Space Bar
153-	اردو شروع	[x x]

(ج) تحریر حروف کی صوتی ترتیب

ا، اِ، أ، آ، و، وھ، ہ، ة، ح، ھ، ء، ع، ی،
یھ، ے، ژ، ب، بھ، پ، پھ، ت، تھ،
ٹ، ٹھ، ط، ج، جھ، چ، چھ، د، دھ، ڈ،
ڈھ، ر، رھ، ڑ، ڑھ، ز، ذ، ض، ظ، س،
ث، ص، ش، ف، ق، ک، کھ، خ،
گ، گھ، غ، ل، لھ، م، مھ، ن، نھ، ں،
ںھ، ںٓ، لا، ـً، ـِ، ـُ، ـٗ، ـٖ، ـٰ،
ـٔ، ـٍ، ـٌ، ـٓ، ـّ،

English -----> Urdu Transliteration
for E-mail addresses

a	=	ا
b	=	ب
c	=	ہ
d	=	د
e	=	ے
f	=	ف
g	=	گ
h	=	ہ
i	=	ـِ
j	=	ج
k	=	ک
l	=	ل
m	=	م
n	=	ن
o	=	ـو (واو مجہول)
p	=	پ
q	=	ق
r	=	ر
s	=	س
t	=	ت
u	=	ـُ
v	=	و
w	=	و
x	=	لا
y	=	ی
z	=	ز

اضافت کونہیں۔ چنانچہ یہاں ہر دو اضافت کے لیے ب''آئ'' کی پہلے سے مستعمل صورت ہی کوملحوظ رکھا گیا ہے۔موقع محل کے مطابق سافٹ ویئر انجن اسے ہمزہ یا زیر میں تبدیل کر دے گا۔

ہر لفظ کے مکمل ہونے پر کی علامت درکار تھی جو Space Bar پوری کرتا ہے تا کہ کسی مغالطے یا لفظوں کے باہم ادغام کا مسئلہ پیدا نہ ہو۔اسی طرح جب کمپیوٹر پر اُردو متن شروع کرنا ہو تو اسے ہدایت دی جائے کہ اب اُردو متن شروع ہو رہا ہے۔اس کے لیے [xx] کی علامت تجویز کی گئی ہے۔

مقتدرہ قومی زبان نے شعبہ اُردو اطلاعیات کے قیام سے قبل ۲۰ اگست ۱۹۹۹ء کو انٹرنیٹ اور ای میل پر اُردو کے استعمال کے امور میں مشاورت کے لیے ایک ذیلی کمیٹی بنانے کا اعلان کیا، جس کے کنویز جناب ظہیر عزیز تھے اور ارکان میں ڈاکٹر خالد رشید، ڈاکٹر محمد افضل، جناب طارق حمید، جناب محمد خاور، جناب طاہر مفتی، جناب این آر لیوال شامل تھے۔شعبہ کے قیام (۲۰ ستمبر ۱۹۹۹ء) کے بعد بھی یہی ذیلی کمیٹی کام کرتی رہی۔

۲۲ جولائی ۲۰۰۰ء کو اُردو معیار بندی میں مشاورت کے لیے قائم کردہ ٹیکنیکل کمیٹی کے اجلاس میں اس ذیلی کمیٹی کی رپورٹ نہ ہونے کے برابر تھی۔اس روز طے ہوا کہ اگلے اجلاس میں یہ ذیلی کمیٹی بھی اپنی رپورٹ پیش کرے۔اس ذیلی کمیٹی کے ۲۔ستمبر ۲۰۰۰ء کے اجلاس میں جن ضروریات کا ذکر ہوا ان میں یہ بھی تھا کہ اُردو میں انٹرنیٹ ویب سائٹ کے لیے نقل حرفی کی معیاری فہرست تیار کی جائے۔چنانچہ ایک تجویز ۹ ستمبر ۲۰۰۰ء کے اجلاس میں پیش کی گئی۔اجلاس میں جناب ظہر عزیز، جناب طاہر مفتی، جناب طارق حمید، میجر سہیل اقبال، راقم اور جناب اکبر سجاد شریک تھے۔راقم کے اس مجوزہ نظام پر غور رخوض شروع ہوا۔کئی منازل گزارنے کے بعد ۱۵ ستمبر ۲۰۰۰ء کو طے ہوا کہ:

۱۔متبادلات کا یہ نظام قابل تحسین ہے اور کام کر سکتا ہے۔

۲۔ اضافی علامات کے بارے میں بار بار چیک کر کے معلوم کر لیا جائے کہ کہیں یہ اُردو ویب سائٹ یا ای میل میں کسی اور کنٹرول مفہوم میں تو استعمال نہیں ہوتیں۔

۳۔ یہ بھی معلوم کیا جائے کہ ویب اور ای میل پتوں کے علاوہ سرچ انجن ان کے ذریعے کلیدی الفاظ اور متن بھی تلاش کر سکے گا۔نیز رومن میں لکھے ہوئے پتوں کو بھی پڑھ سکے گا یا نہیں۔

۴۔ یہ طے کر لیا جائے کہ انگریزی میں متن پہلے سے موجود پتوں کے اُردو متبادل کس نظام اور تحدید کے ساتھ اُردو میں لکھے جائیں گے۔

چنانچہ مختلف مشوروں کے بعد اس نظام میں پانچ بار تبدیلی کی گئی اور بالآخر ۱۷ اکتوبر ۲۰۰۰ء کو یہ طے پایا کہ اسے ''حتمی صورت'' میں (Final Version) کے طور پر شائع کر دیا جائے اور اس کے ذریعے اُردو کے بنیادی الفاظ کی ایک ڈکشنری بھی بنا کر مقتدرہ کے ویب سائٹ پر رکھ دی جائے تا کہ سافٹ ویئر بنانے والے وہاں سے ذخیرہ ڈاؤن لوڈ کر سکیں۔اسے کتابی صورت میں بھی شائع کر دیا جائے۔

اُردو اطلاعیات اور فرہنگ

قوموں کی زندگی میں کبھی کبھار ایسے مراحل اور مواقع بھی آتے ہیں جنھیں تاریخ ساز اور عہد آفرین کہا جا سکتا ہے۔ اُردو پاکستان کی قومی زبان ہے اور اس حقیقت سے انکار نہیں کیا جا سکتا کہ پچھلے ساٹھ برس میں اُردو کو بحیثیت قومی زبان اس کا جائز مقام حاصل نہیں ہو سکا۔ اس اثنا میں بھی زمانے کی تیز رفتار ترقی کے ساتھ ساتھ اُردو کو ساتھ لے کر چلنے کی کچھ اکا دکا اور منتشر کوششیں بھی کی جاتی رہی ہیں۔ یہ کوششیں کبھی عوامی سطح پر ہوئیں اور کبھی ان کا سہرا محبِّ وطن قیادت کے سر جاتا ہے۔ ۱۹۷۹ء میں اپنے قیام سے لے کر اب تک مقتدرہ قومی زبان بھی اس کام میں پیچھے نہیں رہا۔ دفتری امور کو دستی طور پر انجام دینے کا دور ہو یا اُردو ٹائپ رائٹر کی تیاری کے لیے تکنیکی پہلوؤں سے معیاری کلیدی تختہ یا پھر اُردو تار گھر اور ٹیلی پرنٹر کے لیے کلیدی تختہ کی تیاری، غرضیکہ مقتدرہ قومی زبان نے ہر موقع پر قومی خدمت کے جذبے کے ساتھ بڑھ چڑھ کر حصہ لیا ہے۔

دنیا اطلاعات و معلومات کی شاہراہ پر گامزن ہوئی تو بیسویں صدی کی آخری دہائی کے ماہرین لسانیات اور ماہرین کمپیوٹر ٹیکنالوجی کے تمام عالمی ماہرین زبانوں کے بارے میں یک زبان تھے کہ اکیسویں صدی میں صرف وہی زبان زندہ رہ سکے گی جو کمپیوٹر کی زبان بن جائے گی۔ کمپیوٹر کی زبان کے عام طور پر یہ معنی لیے جاتے ہیں کہ زبان کو کمپیوٹر پر ٹائپ رائٹر کی طرح کمپوز کیا جا سکے۔ عام آدمی حتیٰ کہ پڑھا لکھا طبقہ بھی یہی سمجھتا ہے کہ وہ اپنی روزمرہ کی ضروریات، خطوط اور رسائل کی کتابت کا کام اس مشین سے لے۔ مگر وقت نے ثابت کیا کہ جب یہ دنیا ایک گلوبل ویلج بن گئی اور کسی ایک کمرے میں بیٹھ کر پوری دنیا کے تمام علوم وفنون پر تحقیق اور تدریس کا کام بھی اس مشین پر لیا جانے لگا تو معلوم ہوا کہ یہ مشین دنیا بھر کے علوم وفنون کا مخزن ہے اور حصول معلومات، ترجمہ کاری اور لفظ کاری کے لیے اس مشین پر اپنی زبان کا ہونا بے حد ضروری ہے۔ چنانچہ مقتدرہ قومی زبان نے بھی عالمی معیارات کے مطابق کمپیوٹر کی سکرین کو اُردو میں مکمل طور پر بدلنے کا بیڑا اٹھایا۔

اس موقع پر مائیکروسافٹ جیسے بین الاقوامی ادارے کو بھی اس حقیقت کا اندازہ ہو گیا تھا کہ اب لسانی توسیع میں بھی بہت سا کاروبار پوشیدہ ہے، چنانچہ دوسری زبانوں کی طرف بھی توجہ دی جانی چاہیئے یا

انھوں نے اکیسویں صدی کے آغاز ہی میں اپنی تمام تر مصنوعات کو دنیا کی منتخب کردہ سوز بانوں میں پیش کرنے کا فیصلہ کیا۔ انھی سوز بانوں میں سے ایک اُردو ہے جو اب عربی کے بعد ان پہلی چار زبانوں میں شامل ہے جن میں مائیکروسافٹ کی مصنوعات پیش کی جا رہی ہیں چنانچہ کمپیوٹر سکرین کو اُردو میں بدلنے کے سلسلے میں مائیکروسافٹ اور مقتدرہ قومی زبان کے درمیان ایک دوستانہ معاہدہ ہوا۔ جس کی رو سے مائیکروسافٹ کی مصنوعات/ ایپلی کیشن اُردو میں ڈھالنے کا آغاز ہوا۔

اس بین الاقوامی ادارہ نے اپنی تمام مصنوعات مائیکروسافٹ آفس (Microsoft Word, Excel, PowerPoing, PowerPoint, Access, Outlook Express) اور آپریٹنگ سسٹم مائیکروسافٹ ونڈوز XP میں موجود تمام ہدایات، مینیو، دریچے اور کمپیوٹر سکرین کو اُردو میں بدلنے کے کام کا آغاز مقتدرہ قومی زبان کے ساتھ اسی معاہدہ کے تحت کیا۔ مذکورہ بالا تمام مصنوعات کو اُردو میں بدلنے کے لیے انگریزی میں استعمال ہونے والے تقریباً ساڑھے آٹھ لاکھ الفاظ اور اصطلاحات پر مشتمل مواد کے معیاری ترجمہ، کوائف کاری اور تکنیکی تدوین، نظرِ ثانی و معیار بندی کا کام شامل تھا۔ اس منصوبے کی اہمیت اور تکنیکی نوعیت کا اندازہ اس بات سے لگایا جا سکتا ہے:

الف۔ تمام ترجمہ شدہ مواد مائیکروسافٹ آفس کے لیے ان کے کوائفیہ یا ڈیٹا بیس Helium میں کوائف کاری، و تدوین و معیار بندی کے بعد ان کے حوالے کیا گیا جبکہ مائیکروسافٹ ونڈوز Xp کا کام اس سے بھی زیادہ تکنیکی نوعیت کا ہے۔ اس میں تمام آپریٹنگ سسٹم کے پیغامات ان کے مختلف کوائفیوں کے ذریعے کو سکرین پر نمودار ہونے والے بٹنوں تک پر اُردو لگا کر بھجوانا تھا۔

ب۔ اس سارے منصوبے میں ہم معنی الفاظ و اصطلاحات کے لیے علیحدہ علیحدہ اُردو مترادفات استعمال کیے جانے تھے، کیونکہ کمپیوٹر پر ایک لفظ یا اصطلاح کا صرف ایک معنی یا مترادف بھی دیا جا سکتا ہے، وجہ ظاہر ہے کہ ہر لفظ اپنا الگ پروگرام یا عملیہ رکھتا ہے۔ ہم معنی الفاظ کا مترادف استعمال ہونے پر معیار بندی ہی متاثر نہیں ہوتی بلکہ پراسیسنگ بھی غلط ہو جاتی ہے۔

۲۔ اس سلسلے میں مذکورہ منصوبے کے پہلے حصے (مائیکروسافٹ آفس) کے نو پراجیکٹوں کے تراجم اور ابتدائی کوائف کاری کو معیار بندی کے بعد ۳۰۔ جون ۲۰۰۴ء کو مائیکروسافٹ کے حوالے کر دیا گیا۔

۳۔ پہلے حصے کے کام پر نظرِ ثانی کے مطابق ترمیم اور حتمی معیار بندی کا کام جاری رہا اور اس کی تکمیل کے ساتھ ہی پراجیکٹ کے دوسرے حصے مائیکروسافٹ ونڈوز XP کے تراجم، کوائف کاری اور دوسرے تمام مراحل کا کام شروع ہوا۔

منصوبے کے آغاز سے پہلے ابتدائی طور پر ۲۵۰۰ بنیادی اصطلاحات (Core Terms) کے ترجمہ کی معیار بندی کا کام کیا گیا تھا۔ ابتدائی اُردو ترجمہ کے بعد ملک بھر سے ماہرین کی ایک منتخبہ کمیٹی میں

ایک ایک اصطلاح پر سیر حاصل بحث کے بعد حتمی ترجمہ کا فیصلہ کیا گیا تھا۔ اس کمیٹی کے دو دو روزہ دو اجلاس منعقد ہوئے۔ تفصیل حسب ذیل ہے:

۱۔ مائیکروسافٹ کی Core Terms کے اُردو ترجمہ کو حتمی شکل دینے کے لیے مترجمین کمیٹی کا اجلاس مقتدرہ کے کانفرنس ہال میں ۱۸ اور ۱۹ اکتوبر ۲۰۰۳ء (دو روزہ) منعقد ہوا۔ اجلاس میں درج ذیل اراکین نے شرکت کی:

۱۔ جناب سعید احمد (PCB)، اسلام آباد
۲۔ ڈاکٹر محمد افضل (Barani University)، راولپنڈی
۳۔ ڈاکٹر سرمد حسین (FAST)، لاہور
۴۔ جناب علیم احمد، کراچی (نمائندہ جناب ہمایوں قریشی)
۵۔ پروفیسر خورشید احمد رضوی، لاہور
۶۔ جناب تفسیر احمد، اسلام آباد
۷۔ ڈاکٹر انجم حمید (مقتدرہ)
۸۔ ڈاکٹر عطش درانی (مقتدرہ)

دو روزہ مسلسل کام کے بعد مذکورہ پراجیکٹ کا صرف ۲۵% تکمیل کے مراحل تک پہنچ سکا۔ باقی کام کو حتمی شکل دینے کے لیے تمام ممبران نے تجویز دی کہ رمضان المبارک کے بعد ایک اور اجلاس منعقد کیا جائے۔

۲۔ حسب پروگرام مائیکرسافٹ Core Terms کی معیار بندی کے باقی کام کی تکمیل کے لیے ۱۳۔۱۴ دسمبر ۲۰۰۴ء کو مترجمین کمیٹی کا دوسرا اجلاس مقتدرہ کے کانفرنس ہال میں منعقد ہوا۔ درج ذیل اراکین نے شرکت کی۔

۱۔ جناب سعید احمد، اسلام آباد
۲۔ ڈاکٹر محمد افضل، راولپنڈی
۳۔ ڈاکٹر خاور ضیا (Beacon House)ء، لاہور
۴۔ جناب ہمایوں قریشی، کراچی
۵۔ ڈاکٹر نذیر سانگی (AIOU)، اسلام آباد
۶۔ جناب علیم احمد، کراچی
۷۔ ڈاکٹر انجم حمید (مقتدرہ)

۸۔ڈاکٹر عطش درانی (مقتدرہ)

کمیٹی نے مذکورہ اصطلاحات کے اُردو ترجمے کا لفظ بہ لفظ جائزہ لیا۔ دو روزہ مسلسل محنت کے بعد ایک ایک لفظ پر بحث کی گئی اور ترجمے کے اس کام کو حتمی شکل دی گئی۔

ابتدائی ۲۵۰۰ اصطلاحات کا یہ ترجمہ مائیکروسافٹ کو بھجوایا گیا چنانچہ آفس اور ونڈوز کے سافٹ ویئر ترجمہ کرنے کے لیے مقتدرہ کے سپرد ہوئے۔ معاہدہ کے آغاز میں مائیکروسافٹ کے نمائندہ نے مقتدرہ کے دفتر میں آ کر مقتدرہ کے عملہ کو منصوبے پر عمل درآمد کے لیے Helium اور دوسرے کوائفیوں میں تربیت دی۔ اس تربیت کی روشنی میں منصوبے کی بروقت تکمیل کے لیے تمام کام کا جائزہ لینے اور لائحہ عمل طے کرنے کے لیے ملک بھر سے ماہرین اور مترجمین کی کمیٹی کا ایک اجلاس ۱۶۔۱۷ فروری ۲۰۰۴ء کو مقتدرہ کے دفتر میں منعقد ہوا۔ اس اجلاس کی رپورٹ درج ذیل ہے:

(الف): مائیکروسافٹ کے لیے اُردو تراجم کے کام کا جائزہ لینے، تراجم کے اصول وضع کرنے اور تقسیم کار کے اصول اور مزید مترجمین کی تلاش اور رسائی کے حصول کے لیے کمیٹی کے سابقہ اجلاس مورخہ ۲۶۔ جنوری ۲۰۰۴ء کی روشنی میں یہ طے کیا گیا کہ کمیٹی کے تجویز کردہ اور چند مزید افراد کو تلاش کر کے شامل کیا جائے اور کام تفویض کرنے کی بنیادیں وضع کی جائیں۔ چنانچہ یا اجلاس ۱۶، ۱۷ فروری ۲۰۰۴ء کو منعقد ہوا۔ اس اجلاس میں مندرجہ ذیل مندوبین نے شرکت کی۔

۱۔ جناب سعید احمد، پی سی بی، اسلام آباد۔
۲۔ جناب خادم علی ہاشمی، ملتان۔
۳۔ جناب اے ڈی میکن، اسلام آباد۔
۴۔ جناب الیاس قدیر طاہر، اسلام آباد۔
۵۔ جناب علیم احمد، کراچی۔
۶۔ جناب سید محمد اویس احمد، اسلام آباد۔
۷۔ محمد ظہیر عزیز، اسلام آباد۔
۸۔ جناب طارق حمید، اسلام آباد۔
۹۔ جناب صفدر رشید، اسلام آباد۔
۱۰۔ جناب بلال احمد، اسلام آباد۔
۱۱۔ جناب محمد سہیل عمر، لاہور۔
۱۲۔ ڈاکٹر اللہ رکھا ساغر، اسلام آباد۔

۱۳۔ جناب شاہد مسعود، معتمد، مقتدرہ قومی زبان۔

۱۴۔ سید عقیل عباس جعفری، مقتدرہ قومی زبان۔

۱۵۔ محمد اکبر سجاد، مقتدرہ قومی زبان۔

۱۶۔ ڈاکٹر عطش درانی (ناظم منصوبہ)۔

اجلاس میں فیصلہ کیا گیا کہ منصوبے کو دو ذیلی گروہوں میں تقسیم کر دیا جائے۔

۱۔ گروہ مترجمین۔ یہ گروہ کاغذ پر ترجمہ، نظر ثانی اور معیار بندی کا کام سرانجام دے گا۔

۲۔ فنی گروہ۔ یہ گروہ موصولہ ترجمہ کی ویلی ڈیشن، ڈیٹا پروسیسنگ، فائل ٹرانسفر اور مائیکرو سافٹ کے معیار کے مطابق کام کی تکمیل کے لیے کام کرے گی۔

دوسرے روز مترجمین نے عملی طور پر ترجمہ کر کے اس بات کا جائزہ لیا کہ اوسطاً کتنا کام روزانہ مکمل کیا جائے کہ منصوبہ بر وقت مکمل ہوسکتا ہے اور اس سلسلے میں کبھی کبھی الجھنیں اور مسائل بھی درپیش ہو سکتے ہیں، ان کے حل کے لیے کوئی پلیٹ فارم وضع کیا جائے۔ ترجمہ، نظر ثانی اور معیار بندی کے لیے بھی وقت دورانیے اور میقات کا تعین کیا گیا:

۱۔ معیار بندی اور کوائف کاری کا تمام کام مقتدرہ کے دفتر میں انجام دیا جائے گا۔

۲۔ ترجمہ کے لیے بنیادی لغت میں توسیع کی جائے گی اور اسی کو معیار بندی کا آلہ بنایا جائے گا۔

۳۔ دوران ترجمہ میں کوشش کی جائے گی کہ اُردو ترجمہ بھی اتنے ہی حروف (Characters) پر مشتمل ہو، جتنے حروف انگریزی متن میں موجود ہیں۔

۴۔ دوران ترجمہ میں مخففات (Abbreviations)، ٹریڈ مارک، اور قوسین میں ٹیگ کا ترجمہ نہیں کیا جائے گا۔ بلکہ ایسے الفاظ کو انگریزی ہی میں رہنے دیا جائے گا۔

۵۔ کوشش کی جائے گی کہ کوئی لفظ دس حروف سے زیادہ نہ ہو۔

۶۔ ہدایات کا ترجمہ کرتے وقت درمیانی راستہ اختیار کیا جائے گا۔ مثلاً دیکھو، دیکھیں اور دیکھیے میں سے "دیکھیں"۔ اسی طرح رکھو، رکھیں، رکھیے میں سے "رکھیں"۔

۷۔ فعل یا مصدر بنانے کے لیے کرنا یا ہونا یا صیغہ امر کے لیے کریں، بنائیں وغیرہ کا اضافہ کیا جا سکتا ہے۔

۸۔ "and" کے ترجمے کے لیے کوما، و، اور وغیرہ میں سے "و" کو ترجیح دی جائے گی۔

اس سلسلے میں جن تکنیکی و لسانی نوعیت کے مسائل اور مشکلات کا سامنا کرنا پڑا وہ درج ذیل ہیں:

۱۔ مائیکرو سافٹ کی طرف سے سب سے پہلا مطالبہ یہ تھا کہ کسی بھی انگریزی اصطلاح کے لیے اُردو

میں متبادل ایک ہی اصطلاح مختص کی جائے اور کسی بھی ہم معنی دوسرے لفظ کی جگہ وہ لفظ استعمال نہ کیا جائے۔تقریباً ساڑھے آٹھ لاکھ الفاظ میں اس معیار کو برقرار رکھنا انتہائی مشکل کام تھا۔

۲۔ دوسری شرط یہ تھی کہ کسی بھی فقرے یا اصطلاح کا اُردو ترجمہ انگریزی اصطلاح سے حجم میں صرف ۱۰٪ تک زائد ہو۔اس سے زائد ہونے کی صورت میں کوائفیہ اس فقرے کی کوائف کاری کو ناممکن بنا دے گا۔جس کے لیے ضروری تھا کہ ترجمہ حروف کی تعداد کے لحاظ سے کسی بھی صورت ۱۰٪ سے زائد نہ ہو۔

۳۔ اپنے کوائفیہ یا ڈیٹا بیس میں مائیکروسافٹ نے تمام تر اصطلاحات اور فقرات کی پروگرام کاری یوں کر رکھی تھی کہ فقرات اور متن کو تکنیکی ٹیگ کے درمیان رکھا گیا۔کسی ایک بھی ٹیگ کی خرابی کی صورت میں کئی کئی ہفتوں پر محیط کام ضائع ہو جاتا تھا۔جس کی وجہ سے ایک ایک ٹیگ کی ساخت، اس کی تصریح اور دیگر خصوصیات کے مطابق کام مکمل کرنا ضروری تھا۔

۴۔ بار بار تصحیح اور ترمیم کے بعد اندازہ ہوا کہ ابھی لاکھوں اصطلاحات ایسی ہیں جن میں متبادل الفاظ استعمال ہوئے ہیں جنھیں انسانی ذہن کی رسائی سے باہر سمجھا جا سکتا ہے۔لیکن مائیکروسافٹ کا میکانکی آلہ Consistancy Checker ایک ایک لفظ اور فقرے پر رک جاتا اور جب تک اس کی تصحیح نہ کر لی جاتی آگے بڑھنا مشکل تھا۔

۵۔ مائیکروسافٹ کی ٹیکنالوجی کو اُردو میں ڈھالنا اس لیے بھی مشکل نظر آتا تھا کہ یہ انگریزی کی طرح بائیں سے دائیں انداز تحریر کی بجائے دائیں سے بائیں کا انداز تحریر رکھتی ہے۔پھر اُردو جملے کے اندر انگریزی یا رومن حروف میں بعض آلات اور ٹریڈ مارک کو قائم رکھنا بھی مشکل امر تھا۔

۶۔ ان تراجم کی بنیاد پر آزمایشی طور پر سافٹ ویئر (Microsoft Office 2003) تیار کر کے اُردو Build حتمی جائزے اور تصحیح کے لیے بھی تقریباً چار بار مقتدرہ کے حوالے کیے گئے جن ایک ایک کمانڈ، ہدایت، مکالمہ خانوں، ان کے ذیلی مکالمہ خانوں تک ایک ایک لفظ کا بغور جائزہ لیا گیا۔ اس مرحلے پر بھی مقتدرہ میں تمام تیار سافٹ ویئر میں ہر کمانڈ اور اس کی فعالیت کی بنیاد پر بھی تراجم اور معیار بندی کی گئی۔

مائیکروسافٹ آفس 2003ء کے تکمیلی مراحل کے اختتام سے پہلے ہی Microsoft Windows Xp کے ترجمے کا کام بھی شروع کر دیا گیا۔اس منصوبہ پر عمل درآمد کا آغاز Microsoft Office سافٹ ویئر کے ترجمے میں درپیش تمام تر تجربات اور مشکلات اور ان کے حل کو سامنے رکھتے ہوئے کیا گیا۔ اس مرحلے میں کام کے آغاز ہی سے مائیکروسافٹ ٹیم کی طرف سے فروری ۲۰۰۵ء تک تکمیل کا شیڈول دیا گیا۔تراجم کا یہ کام مارچ ۲۰۰۵ء میں مکمل ہوا۔ جن کی MIS اور دیگر تکنیکوں میں مہارت سے

تبادلہ امسلہ ممکن ہوسکا۔ ان میں جناب محمد اسد زمان، جناب محمد نعمان، جناب محسن حجازی اور جناب نوازش علی کا ذکر لازم ہے۔

مائیکروسافٹ کے آفس اور ونڈوز کی اُردو زبان میں منتقلی جیسے اہم ترین منصوبے پر کام شروع کرنے کے ساتھ ہی حکومت پاکستان کی طرف سے "مرکزِ فضیلت برائے اُردو اطلاعیات" کے ایک بہت بڑے منصوبے کے پہلے مرحلے کے پی سی۔ا کی منظوری مرحمت فرمادی تا کہ اُردو کی کمپیوٹر کاری کے حوالے سے ترقی کی تمام تر کوششوں کو یک جا کیا جا سکے اور ایک قومی پلیٹ فارم مہیا ہو۔ سیکرٹری کابینہ جناب اعجاز رحیم کی ذاتی دلچسپی اور تعاون سے ان کاموں کا آغاز اور تکمیل ممکن ہوئی۔ اس منصوبے کا انتظامی آغاز یکم جولائی ۲۰۰۴ء سے ہوا۔ اس میں ابتدائی طور پر یونی کوڈ کی بنیاد پر نستعلیق کا حروف پر مشتمل فانٹ کی تیاری، انگریزی سے اُردو مشین ٹرانسلیشن سافٹ ویئر کی تیاری کے لیے کمپیوٹیشنل گرامر کی تیاری اور دیگر مراحل کی تکمیل شامل ہے۔

مرکزِ فضیلت کے منصوبے کے اہم نکات:

الف: اُردو اطلاعیات کے مرکزِ فضیلت کا قیام

(تین مرحلوں میں)

پہلا مرحلہ: موجودہ پی سی۔I

۱۔ معیاری نستعلیق فانٹ کی تیاری (جو یونی کوڈ پر مبنی ہو)

۲۔ مشینی ترجمے کے لیے کمپیوٹر کے حوالے سے اُردو قواعد لکھنا۔

۳۔ مشینی ترجمہ (مرحلہ اول) کا معیاری سافٹ ویئر تشکیل دینا۔

ب: جامعات کو اُردو اطلاعیات کا شعبہ قائم کرنے میں مدد دینا:

(خاص طور پر وفاقی اُردو یونیورسٹی کے حوالے سے پائلٹ پروگرام جاری کرنا)

۱۔ نصاب۔

۲۔ ٹیوٹوریل وغیرہ۔

۳۔ تحقیقی منصوبوں کی فہرست۔

ج: حکومت پاکستان کو دفتروں، عدالتوں اور تعلیمی اداروں میں اُردو زبان کے انتخاب کا حق دینے کے لیے تکنیکی مدد فراہم کرنا۔

د: مرکزِ فضیلت کے حوالے سے معیارات کی تشکیل کے لیے ایسے تمام تحقیقی امور انجام

دینا جو اُردو زبان کے حوالے سے اس مرکز کو حتمی استناد کا درجہ دے سکیں۔

حروف کی بنیاد پر معیاری نستعلیق فانٹ

اس وقت تک تمام موجود سافٹ ویئر حروف کی بجائے شکل کی بنیاد پر تشکیل دیے گئے ہیں۔جس کی وجہ سے حرفی صورت میں بننے والے فانٹ سے ۱۰۰ گنا زیادہ جگہ گھیرتے ہیں۔یہی وجہ ہے کہ ویب سائٹ پر اُردو تصویری بنیاد پر ہوتی ہے اور ویب سائٹ کھلنے میں بہت زیادہ وقت لگتا ہے لہذا اسے ناکام تصور کیا جاتا ہے۔اس حرفی بنیاد پر معیاری اُردو نستعلیق فانٹ کی تیاری کے بعد اُردو بھی انگریزی اور دیگر زبانوں کے برابر جگہ لے گی اور نیٹ پر آسانی سے سفر کرنے کے قابل ہو جائے گی۔مائیکروسافٹ جیسا ادارہ بھی کسی ایسے اُردو فانٹ کو زیر استعمال لانے میں مشکل محسوس کرتا ہے جو GIF کی بنیاد پر کام کرے۔اُردو آفس اور ونڈوز کے لیے بھی ایک ایسے معیاری نستعلیق فانٹ کی ضرورت ہے جو بٹ کی بنیاد پر اور یونی کوڈ کی معاونت سے کام کر سکے۔مقتدرہ نے اسی خاطر یونی کوڈ جیسے عالمی ادارے کی رکنیت حاصل کر رکھی ہے۔جس کے ورژن ۴۔۰۰ میں بعض حروف اور علامات مقتدرہ کی سفارشات کی روشنی میں شامل کی گئی ہیں۔اب نستعلیق فانٹ کے بعد بھی بعض علامات اور نقاط پر یونی کوڈ کی معاونت درکار ہوگی۔

کمپیوٹیشنل گرامر

دنیا بھر میں متعدد کوششوں کے باوجود انگریزی سے اُردو میں مشین ٹرانسلیشن کے سافٹ ویئر بنانے کی راہ میں سب سے بڑی رکاوٹ کسی معیاری اُردو کمپیوٹیشنل گرامر کی عدم دستیابی ہے۔اُردو چونکہ ایک موضوعی زبان ہے اس لیے اس کے تکنیکی قواعد انگریزی کے قواعد کے ضابطوں پر منطبق کرنا بے حد مشکل ہے اور تقریباً ناقابل عمل کام ہے۔اس لیے کمپیوٹر پر مشینی ترجمہ کاری سے قبل ضروری ہے کہ کوئی ماہر لسانیات جدید ٹیکنالوجی کی ضروریات کو مدنظر رکھتے ہوئے ایک کمپیوٹیشنل گرامر تحریر کرے تاکہ اسے بنیاد بنا کر ماہرین کمپیوٹر کے اُردو میں آمدہ مسائل حل کر سکیں۔

مشین ٹرانسلیشن (مشینی ترجمہ) (انگریزی، اُردو)

مشینی ترجمہ کی اہمیت بین الاقوامی سطح پر ایک مسلمہ حقیقت کی شکل اختیار کر چکی ہے۔اس لیے جب تک انگریزی سے اُردو میں مشینی ترجمہ کی روایت کا آغاز نہیں ہو جاتا، اس وقت تک جدید علوم کا اُردو میں فوری حصول ناممکن ہے۔اس لیے اس اہم ترین ضرورت کو بھی مدنظر رکھتے ہوئے اس پراجیکٹ میں دیگر معیارات

کے ساتھ مشینی ترجمہ کا منصوبہ بھی رکھا گیا ہے جو کم از کم دفتری ضروریات کو تکنیکی سطح پر پورا کر سکے۔

ستمبر ۲۰۰۵ء سے اُردو آفس اور ونڈوز کے اجراء کے موقع پر اس بنیادی فرہنگ کی اشاعت مقتدرہ کی طرف سے کی گئی جو معیار بندی کے لیے بنیاد کا کام دے رہی ہے اور دیگر کمپیوٹر ایپلی کیشنز کی اُردو میں تیاری کے لیے بھی ایک عالمی ذخیرہ اور معیار کے طور پر استعمال ہو رہا ہے۔ یقیناً یہ فرہنگ اپنی نوعیت کی پہلی کاوش ہے۔ عالمی سطح پر اُردو کے تکنیکی فروغ اور ترقی کے لیے یہ فرہنگ ناگزیر حیثیت رکھتی ہے۔

("اخبارِ اُردو"، اسلام آباد، ستمبر ۲۰۰۵ء)

۶

کمپیوٹر کے لیے اُردو الفاظ سازی
(بنیادی مسائل)

اُردو میں مائیکروسافٹ کی ایپلی کیشن خاص طور پر آفس ۲۰۰۳ء اور ونڈوز ایکس پی/ سٹارٹر ایڈیشن کو اُردو میں منتقل کرنے اور یوں کمپیوٹر سکرین کو اُردو کی صورت دینے کے دوران میں کئی قسم کی مشکلات اور مسائل سے واسطہ پڑا۔ جن کا لب لباب کچھ یوں ہیں:

اُردو کے بارے میں بعض مقولات بنیادی طور پر غلط فہمی اور خوش فہمی پر مبنی ہیں۔ مثلاً یہ کہ:

۱۔ **اُردو کوئی لشکری یا کھچڑی سی زبان ہے۔ اس کے کوئی با قاعدہ قواعد وضوابط نہیں اور ہر طرح کی آمیزش بلا کسی تردد کے اس میں کی جا سکتی ہے،** ایسا نہیں۔ اُردو ایک باقاعدہ زبان ہے جسے اس دور کی اعلیٰ سول سوسائٹی نے تشکیل دیا جب یہ زبان پروان چڑھ رہی تھی۔ اس کے قواعد و ضوابط باقاعدہ طور پر موجود ہیں۔ الفاظ سازی کے اصول طے شدہ ہیں۔ یہ الگ بات کہ ہمارے اکثر اُردو دان ان سے آشنا نہیں۔

۲۔ **اُردو میں ابلاغ کی مکمل صلاحیت پہلے سے موجود ہے اور اس میں ہر بات بیان کی جا سکتی ہے۔** یہ بات محض خوش فہمی کا نتیجہ ہے۔ اُردو ایک موضوعی (Subjective) زبان بنا دی گئی ہے۔ اس میں ہر فرد اپنی مرضی استعمال کرتا ہے۔ خاص طور پر تراجم اور تکنیکی ابلاغ کے حوالے سے یہ بات طے شدہ ہے کہ اُردو میں ابلاغ و بیان میں ابتری پائی جاتی ہے۔ چنانچہ مشینی سطح پر دنیا کی دوسری بڑی زبانوں مثلاً انگریزی، فرانسیسی، جرمن، عربی وغیرہ کے ساتھ اس کو موازن نہیں کیا جا سکتا۔ اُردو میں پہلے سے ابلاغ کی ہر سطح کی صلاحیت کیونکر ممکن ہے۔ جب تک کہ کسی خاص میدان کے لوگ اسے اپنے موضوع اور مضمون کے حوالے سے اس کام لیے تیار اور استعمال نہیں کرتے۔

۳۔ **اُردو میں تراجم کو آسان/ عام فہم اور براہ راست ابلاغ کا حامل ہونا چاہیئے۔ بات سمجھ میں آنی چاہیئے۔** یہ غلط فہمی اس پہلو سے قابلِ گرفت ہے کہ ابلاغ کی سطح کا خیال رکھے بغیر اور علمی

گفتگو کو عام آدمی کے فہم کے ساتھ منطبق کر کے یہ کہا جا سکتا ہے کہ اُردو ہر مضمون کے لیے سہل اور آسان بیان پیش کرے۔ یہ آسان کرنے کی فرمایش یعنی آسان بیان کی کوشش محض صحافتی نوعیت کی ہوتی ہے۔ جو کا مقصد معلومات کو عام آدمی تک پہنچانا ہوتا ہے۔ چونکہ عام آدمی کا ذخیرہ الفاظ محدود ہوتا ہے۔ اس لیے ہر قسم کا مضمون اس محدود ذخیرے کے اندر بیان ہو۔ علمی اور تکنیکی سطح پر ایسا ممکن نہیں ہے۔ علمی اور تکنیکی زبان محض خواص کے لیے ہوتی ہے اور اسے انگریزی میں بھی زبان برائے خصوصی مقاصد (LSP) کا نام دیا جاتا ہے۔ انگریزی میں بھی علمی زبان براہ راست سمجھ میں نہیں آ سکتی جب تک کہ اس کی مخصوص اصطلاحات کا مخصوص مفہوم جان نہ لیا جائے۔ صحافتی زبان کا مطلب یہ ہے کہ لفظ عمومی معنی کو بیان کرے جبکہ علمی زبان میں لفظ خصوصی معنی کے لیے استعمال ہوتا ہے۔

اب سب سے بڑا مسئلہ یہ ہے کہ علم کی اکائیاں بنیادی تصورات ہیں۔ دنیا بھر میں ڈھائی سو سے زائد علوم وفنون ہیں، جن کی تصوراتی اکائیوں کو ایک جرمن محقق نے گن کر ان کی تعداد ایک کروڑ ستر لاکھ بتائی ہے۔ علمی دنیا میں یہ ضروری ہے کہ ہر تصور علیحدہ اصطلاح کی صورت میں بیان ہو۔ چنانچہ ایک کروڑ ستر لاکھ (۱۷،۰۰۰،۰۰۰) اصطلاحات درکار ہیں۔ مجبوری یہ ہے کہ دنیا کی کسی بھی زبان میں بنیادی الفاظ زیادہ سے زیادہ ساٹھ ہزار ہیں۔ ادبی دنیا ان میں سے پندرہ سے بیس ہزار الفاظ استعمال کرتی ہے۔ شیکسپیئر نے، جسے دنیا میں زیادہ الفاظ استعمال کرنے والا سب سے بڑا ادیب سمجھا جاتا ہے، کل بائیس ہزار الفاظ استعمال کیے ہیں۔ صحافتی دنیا اپنا کام دس ہزار الفاظ سے چلا لیتی ہے۔ روزانہ بات چیت اور بنیادی ضرورتوں کے لیے پندرہ سو الفاظ کا ذخیرہ کافی ہوتا ہے۔

یونیسکو کی سفارشات کے مطابق کسی بھی زبان میں اگر تین لاکھ اصطلاحات کا ذخیرہ موجود ہو تو وہ علمی مطالب بیان کرنے کے قابل ہوتی ہے۔ لکسمبرگ کے ایک اصطلاحی مرکز (یورپی یونین) کے مطابق انگریزی میں اصطلاحات کا ذخیرہ پانچ لاکھ ساٹھ ہزار ہے۔ علمی مطالب تو کروڑوں ہیں چنانچہ علمی دنیا نے مخففات پر انحصار کرنا شروع کیا تو یہ تعداد ایک لاکھ دس ہزار ہوئی۔ ہر سال انگریزی اصطلاحات میں کوئی دو ہزار کا اضافہ ہوتا رہتا ہے۔ چنانچہ انگریزی زبان مخففات سمیت سات لاکھ کی تعداد کو چھوتی ہے۔ پھر ہر اصطلاح کئی کئی میدانوں میں متعدد معنی کے لیے استعمال کی جاتی ہے۔ جیسے Culture کی اصطلاح ۱۴۸ معنی میں استعمال ہوتی ہے۔ گویا سات لاکھ اصطلاحیں چالیس پچاس لاکھ تصورات کو بیان کرتی ہوں گی۔ اس سے نتیجہ یہ نکلتا ہے کہ ابھی کروڑوں تصورات کے بیان کے لیے انگریزی بھی ناکافی ہے۔ اُردو میں کل ذخیرہ اصطلاحات (قدیم وجدید) ابھی تک دو تا ڈھائی لاکھ ہے جبکہ اُردو کے بنیادی الفاظ کی تعداد صرف بتیس ہزار ہے۔ کیا اسی بل بوتے پر ہم اُردو کو علمی اور تکنیکی زبان بنانے کا عزم رکھتے ہیں؟

مخصوص اور خصوصی مطالب کو بیان کرنے کے لیے مخصوص الفاظ اور خاص زبان کی ضرورت ہوتی ہے۔ خاص زبان صرف ماہرین کی زبان ہوتی ہے جو رفتہ رفتہ چھن چھن کر تھوڑی بہت حد تک عام زبان میں شامل ہوتی رہتی ہے۔ اس لیے یہ توقع کرنا کہ خاص زبان کے تصورات/ مفاہیم عام زبان میں ادا ہو سکیں تقریباً ناممکن ہے۔ خاص زبان ماہرین کے باہمی ابلاغ کے لیے ہوتی ہے، عوام کے ساتھ ابلاغ کے لیے نہیں۔ اس لیے جب کوئی عام فرد کسی خاص موضوع/مضمون/ میدان میں کچھ سیکھنا چاہتا ہے تو اسے اس میدان کے تصورات، مفاہیم اور ان کے لیے وضع کردہ اصطلاحات سے واقفیت حاصل کرنا ہوتی ہے۔

خاص زبان کے لیے اصطلاحات سازی کا کام بھی عمومی لفظ سازی کے اصولوں پر نہیں ہوتا۔ علمی زبان کے الفاظ قرارداد (Convention) کے ذریعے طے کیے جاتے ہیں۔ خاص الفاظ سازی کا یہ حق بھی محض زبان دانوں، ادیبوں اور شاعروں کو نہیں بلکہ اصل کام کرنے والے موجدوں، دریافت کنندوں، استادوں، محققوں، مقالہ نگاروں اور ان مقالہ جات کے خاص مترجموں کو حاصل ہوتے ہیں۔

ان حوالوں سے اگر ہم دیکھیں تو زبان کو آسان بنانے اور عام فہمی کا اصول محض زبان کے نادان دوستوں کا مطالبہ رہ جاتا ہے یا پھر ان لوگوں کا پراپیگنڈہ/ہتھیار ہے جو نہیں چاہتے کہ اُردو زبان علمی طور پر ترقی کرے اور یہ محض ادبی اور صحافتی دنیا تک محدود ہو کر رہ جائے یعنی بس مطلب بیان ہو، تصورات کی تفہیم نہ ہو۔

ترجمہ کاری میں یہ پہلو اور بھی خطرناک صورت اختیار کر لیتا ہے۔ ادبی وصحافتی تراجم سے قطع نظر علمی اور تکنیکی میدانوں میں ترجمے کا کام عام طور پر دو طرح سے انجام پاتا ہے۔ ایک انفرادی مترجمین جو زیادہ تر اپنی مرضی کے تراجم سامنے لاتے ہیں اور اپنے من پسند الفاظ شامل کرتے ہیں۔ دوسرے ادارہ جاتی طریقے جن میں مجالس اور کام کی کئی تہیں ہوتی ہیں اور وہاں بنیادی فرہنگیں بطور معیار، یکسانیت استعمال کی جاتی ہیں۔

ان تراجم میں بھی دو بڑے رجحانات عام ہیں۔ ایک زبان کے تخلیص پسند مترجمین جو ہر مترادف عربی، فارسی ذخیرے سے ڈھونڈتے ہیں یا بہت ہوا تو کوئی مقامی لفظ شامل کر لیں۔ دوسرے لفظی تارید یا اُردوانے کا عمل انجام دینے والے لوگ جو انگریزی اصطلاحوں کو بجنسہ رکھنے کے حامی ہیں۔ دونوں انتہائیں تکنیکی زبان وضع کرنے کے لیے قابل قبول نہیں۔ خاص طور پر کمپیوٹر سائنس میں ان سے کام نہیں چلتا۔ عربی، فارسی خود بھی تعریب اور تفریس پر گزارا کر رہی ہیں اور انگریزی الفاظ کو جوں کا توں لینے میں ان سے دیگر صورتیں اور وضعیں ممکن نہیں۔ ایک امتزاجی صورت ہی سے اس چیلنج سے عہدہ برآ ہوا جا سکتا ہے۔

تکنیکی زبان نہ تو ادبی محضر پر انحصار کر سکتی ہے اور نہ صحافتی انداز اپنا سکتی ہے۔ ادبی زبان میں ایک لفظ میں معنی کے دریا بہائے جا سکتے ہیں اور قاری اپنی مرضی کے مطالب اخذ کرتا ہے جبکہ تکنیکی زبان ایک لفظ کے ایک ہی معنی کو ملحوظ رکھنا چاہتی ہے۔ چونکہ ایسا ممکن نہیں کہ ایک کروڑ ستر لاکھ تصورات کے لیے ایک کروڑ

ستر لاکھ الفاظ وضع ہوسکیں، اس لیے زبان کے بنیادی الفاظ ہی کومختلف میدانوں میں مختلف معنی کے لیے استعمال کرلیا جاتا ہے۔اس عمل کے لیے معیار بندی کا سہارا لیا جاتا ہے۔

معیار بندی صرف لغات اور فرہنگیں مرتب کر دینے کا نام نہیں۔علمی یکسانیت،تصورات اور بیان میں مغالطہ ربائی، درست عملی تفہیم اور تبادلہ علم کے ساتھ ساتھ عالمی تربیت اور کاروباری بہتری کے لیے بھی الفاظ اور اصطلاحات کی معیار بندی ضروری ہوتی ہے۔ یہ معیار بندی لفظوں کے معنی طے کیے یعنی قرارداد کا سہارا لیے بغیر ممکن نہیں ہوتی۔قرارداد کی ایک مثال یہ ہے کہ اگر Cat کے لیے اُردو میں "بلی" کامفہوم ہے تو حیوانیات میں بلی کی قسم کے لیے مختلف جانوروں کو ایک خاندان سمجھنے کے لیے بھی یہی لفظ استعمال کرنا اہل علم کے درمیان علمی تفہیم میں مغالطہ پیدا کرنے کا سبب بنتا ہے۔اب اگر ایک پروفیسر لفظ "بلی" کہے اور دوسرا اس سے "خاندان بلی" مراد لے جبکہ پہلے کا مقصود وہ نہ ہو یعنی Felidae کو "خاندان بلی" کہا جائے جبکہ لازم نہیں کہ اس میں شامل کوئی نوع "بلی" ہو تو دونوں کے درمیان ابلاغ نہ ہو سکے گا۔اسی طرح اگر "dog" کی مثال لی جائے تو dog family کے لیے لاطینی کا سہارا لے کر Canidae کا لفظ استعمال کیا جاتا ہے جبکہ کمپیوٹر سائنس میں dog کا لفظ تیرہ عدد رکھنے والے اعشاری ہندسے کے لیے بھی استعمال ہوتا ہے۔ یہاں "کتے" کا مفہوم بالکل موجود نہیں۔یہ معنی طے کرنے کا یہ طریق کار قرارداد (Convention) کہلاتا ہے یعنی یہاں معانی "طے" کیے گئے ہیں،لفظ میں موجود نہیں۔

مائیکروسافٹ کی کمپیوٹر اپلی کیشنوں کے لیے الفاظ اور جملوں کو ترجمہ کرنے کے دوران میں جن مسائل اور مشکلوں سے واسطہ پڑا ان میں مندرجہ بالا کے علاوہ خاص طور پر مندرجہ ذیل امور شامل تھے:۔

۱۔ عدم یکسانیت، ابتری اور بے قاعدگی۔اس سے بچنے کے لیے بنیادی ترجمہ کاری آلہ "فرہنگ" کی تیاری ضروری تھی۔

۲۔ کوئی بھی جملہ انگریزی جملے سے مقدار میں (حروف کی تعداد) میں دس فی صد سے زیادہ نہ ہو۔

۳۔ انگریزی جملوں کے درمیان آنے والی قوسین۔جن کے اندر کسی پروگرام/ اپلی کیشن کا نام درج ہو، اُردو میں بھی اسی مقام پر آئیں اور ہر لفظ کا ترجمہ دیا جائے۔خاص طور پر حروف The, a, an کا ترجمہ بعض اوقات ممکن نہیں ہوتا کہ اُردو میں شامل کیا جائے مثلاً اس جملے کا ترجمہ ملاحظہ ہو:

The <MS Office> is user friendly programe.

یہ <MS Office> ایک صارف دوست پروگرام ہے۔

اس جملے میں The کا ترجمہ ضروری نہیں تھا،لیکن اگر ایسا نہ کیا جاتا تو قوسین میں دیا ہوا انگریزی لفظ انتہائی بائیں طرف چلا جاتا۔چنانچہ "یہ" کا لفظ اضافہ کیا گیا۔

۴۔ انگریزی میں افعال (Verbs) کے بارہ زمانے اور اُردو کے متعدد افعال (۵۶) میں کس طرح منتقل کیے جائیں اور ان میں سیاق وسباق کے لحاظ سے بہترین جملے کا انتخاب کیسے کیا جائے۔ مثلا اس جملے کے متعدد تراجم ممکن ہیں:

The birds would be flying.

پرندے اڑتے ہوں گے/ پرندے اڑ رہے ہوں گے/ پرندے اڑتے چلے جا رہے ہوں گے۔

سیاق وسباق کے لحاظ سے آخری جملہ درست ہے مگر اس میں انگریزی کے ایک verb کے مقابلے میں اُردو کے پانچ افعال اڑنا، چلنا، جانا، رہنا، ہونا استعمال ہوتے ہیں۔ خود کار کمپیوٹر کیوں کر ایسا کر سکتا ہے۔

افعال کی تذکیر و تانیث، واحد، جمع اور امالہ اور متعلق فعل کی صورت بھی مشکل میں ڈالتی ہے۔ مثلا:

جاتا، جاتی، جاتے، جاتیں، جاتے ہوئے، جاتے جاتے۔

ایک لفظ next کے کئی تراجم "اگلا، اگلے، اگلی، اگلوں، آگے" وغیرہ ممکن ہیں۔ ان میں سے کسی ایک لفظ کا انتخاب کیوں کر ممکن ہے جو ہر قسم کی صورت حال میں استعمال کیا جا سکے۔

۵۔ انگریزی مرکب جملے میں مختلف اجزا اُردو میں اسی نحوی ترکیب کے ساتھ مرقع نہیں ہوتے۔ اس مقصد کے لیے مائیکروسافٹ نے ہمیں جو خصوصی سافٹ ویئر Helium بنا کر دیا، اس سے قدرے سہولت حاصل ہوئی لیکن موقع محل کی مناسبت اور انحصار (Situation Dependency) کے لیے بہت محنت کرنا پڑتی تھی۔

۶۔ اُردو لفظ کی طوالت دس حروف سے زیادہ نہ ہو۔ اسے بھی کسی نہ کسی صورت ملحوظ رکھا گیا۔

۷۔ اُردو میں موضوعیت کو کم سے کم کیا جائے، اس کے لیے ایک معیاری فرہنگ وضع کی گئی۔ بعض امکانی پسند کو محدود کر دیا گیا جیسے دیکھو، دیکھیں، دیکھیے = See

ان میں سے "دیکھیں" کو معیار قرار دیا گیا۔

تلاش، تلاش کریں اور تلاش کرنا = Search

ان میں سے "تلاش" کو بٹن پر اور "تلاش کریں" کو دریچے میں استعمال کیا گیا۔

۸۔ کئی صورتوں میں اسم اور فعل ایک سے ہوتے ہیں، جیسے Searching "تلاش"، "تلاش کرنا"، "تلاش کر رہا ہے"۔ مفہوم کو سمجھنا اور اس کے مطابق مترادف فراہم کرنا ضروری ہوتا ہے۔

کمپیوٹر کی اپیلی کیشنوں کے لیے مترادفات تلاش کرنا آسان کام نہیں۔ ذیل کی مثالوں میں ایسے الفاظ کے گروہ دیے گئے ہیں جو بظاہر یکساں مفہوم رکھتے ہیں اور ان کے لیے ایک سے اُردو مترادفات کا

استعمال ممکن ہے۔مگران میں سے ہر ایک کا وظیفہ (Function) مختلف ہے۔اس لیے ہر ایک کے لیے الگ الگ لفظ درکار ہے۔اس میں مترجم کے ذاتی پسند ناپسند کو بھی دخل ہوسکتا ہےمگر معیار بندی ان کی اجازت نہیں دیتی، اس لیےقرارداد یا متعین الفاظ کا سہارا لینا پڑتا ہے۔اب یہ قرارداد کون کرے، کوئی فرد یا ادارہ؟ لازم ہے کہ ادارہ ورنہ "موضوعیت" کا راستہ نہیں رک سکتا۔

1. Show, Display, Presentation, Exhibition.
2. Find, Search, Explore, Surfing,
3. Location, Place, Space, Memory
4. Point, Dot, Period
5. Event, Situation
6. Top, Over, Upper, Above, On
7. Bottom, Down, Under, Below, Beneath
8. Change, Replace, Convert,
9. Correct, Rectify, Adjust,
10. Exact, Precise
11. Back, Backward, Behind
12. Job, Task, Tasks, Work
13. Advance, High
14. Alert, Warning
15. Allow, Permit
16 Action, Verb
17. Apply, Paste, Connect
18. Complete, Finish, Over
19. Confirm, Certified, Authorised, Verified
20. Custom, Customize, Specify
21. Choose, Select
22. Instrument, Tool, Device
23. Fixing, Correcting, Rectifying

ان میں سے کئی الفاظ ایسے ہیں جن کے معنی لغت میں کچھ اور ہیں مگر کمپیوٹر سائنس نے ان کے لیے اپنے مخصوص معنی متعین کرر کھے ہیں یعنی قرارداد کا عمل کیا ہے۔اُردو میں ان کے مترادفات کچھ یوں وضع کیے گئے۔ملاحظہ ہوں۔

	Word	Lexical Meanings	Covened Meanings
1.	By default	ازنقص/ازروئے خرابی	طے شدہ
2.	Archive	درمحراب/درطاقچہ	محفوظہ
3.	Arguments	دلائل/استدلال	براہین
4.	Bridge	پل	برج
5.	Character	حرف،خوبی،کردار	کریکٹر
6.	Child	بچہ	تابع
7.	Animation	حیوان داری	حرکت اندازی
8.	Browsing	چارا کھانا	براؤز کرنا
9.	Surfing	برموج ہونا	سرفنگ
10.	Show	دکھانا	نمایش کرنا
11.	Display	دکھانا	دکھانا
12.	Presentation	اظہار/پیشکش	پیش کش
13.	Exhibition	نمایش	نمایش
14.	Find	تلاش کرنا	ڈھونڈنا
15.	Search	تلاش کرنا	تلاش کرنا
16.	Explore	تلاش کرنا	ایکسپلور کرنا
17.	Surfing	تلاش کرنا	سرفنگ کرنا
18.	Location	جگہ	جگہ
19.	Place	جگہ	جگہ
20.	Space	جگہ	گنجایش
21.	Memory	یادداشت	میموری
22.	Point	مقام/نقطہ/جگہ	نقطہ/نکتہ/پوائنٹ
23.	Dot	نقطہ	ڈاٹ

24.	Dash	چھوٹا نقطہ	ڈیش
25.	Event	وقوعہ/موقع	وقوعہ
26.	Situation	صورت حال/موقع	موقع محل
27.	Top	اوپر	بالا
28.	Over	اوپر/پر	بر
29.	Above	اوپر	پر
30.	On	پر	بر
31.	Bottom	پینڈا	پایان
32.	Down	نیچے	نیچے/پست
33.	Under	نیچے	تحت
34.	Below	نیچے	نیچے
35.	Beneath	نیچے	نیچے
36.	Change	بدلنا	تبدیل کرنا
37.	Replace	بدلنا	بدلیں
38.	Convert	بدلنا	ادل بدل کرنا
39.	Correct	صحیح کرنا	صحیح کرنا
40.	Rectify	صحیح کرنا	تصحیح کرنا
41.	Fix	صحیح کرنا	درست کرنا
42.	Adjust	صحیح کرنا	بمطابق بنانا
43.	Exact	صحیح	درست
44.	Precise	صحیح	عین
45.	Back	پیچھے	واپس
46.	Backward	پیچھے کو/پسماندہ	پیچھے
47.	Behind	پیچھے	عقب

48.	Work	کام	کار
49.	Task	کام	کام
50.	Tasks	کام	مجموعہ کار
51.	Advance	اعلیٰ	ایڈوانس/اعلیٰ
52.	High	اعلیٰ	اعلیٰ
53.	Alert	ہشیار	انتباہ
54.	Warning	تنبیہ	تنبیہ
55.	Allow	اجازت دینا	اجازت دینا
56.	Permit	اجازت دینا	اجازت دینا
57.	Action	فعل	عمل
58.	Verb	فعل	فعل
59.	Apply	جوڑیں	لگائیں
60.	Paste	جوڑنا/لگانا	جوڑنا
61.	Connect	جوڑنا	جڑنا
62.	Complete	مکمل کرنا	مکمل کرنا
63.	Finish	ختم کرنا	تکمیل کرنا
64.	Over	مکمل/تمام	ختم
65.	Confirm	تصدیق کرنا	توثیق کرنا
66.	Certified	تصدیق شدہ	تصدیق شدہ
67.	Authorised	تصدیق شدہ	مجاز
68.	Verified	تصدیق شدہ	توثیق شدہ
69.	Custom	مخصوص کرنا	مخصوص
70.	Customise	مخصوص بنانا	اختصاص کرنا
71.	Specify	مخصوص بنانا	تخصیص کرنا

72.	Choose	منتخب کرنا	انتخاب کرنا
73.	Select	منتخب کرنا	منتخب کرنا
74.	Instrument	اوزار	آلہ
75.	Tool	اوزار	ٹول
76.	Device	اوزار	آلہ
77.	Fixing	درست کرنا	درست کرنا
78.	Correcting	درست کرنا	تصحیح کرنا
79.	Rectifying	درست کرنا	صحیح کرنا

ان میں سے بعض متعین کردہ اُردو مترادفات سے مجھے بھی ذاتی طور پر اختلاف ہے۔ لیکن جب معیار بندی کا کام اداروں کے کرنے کا ٹھہرا تو میری ذاتی رائے خواہ کچھ بھی ہو اصولی طور پر میں اس تکنیکی مجلس کی وضع کردہ فرہنگ سے متفق ہوں۔ یہی طریق کار معیار بندی کے عالمی اداروں Unicode،ISO وغیرہ میں برتا جاتا ہے۔ جب تک میری رائے یہ تکنیکی مجلس منظور کر کے اگلے ایڈیشن میں شامل نہیں کر لیتی۔ مجھے عملاً اسی کی منظوری سے متفق رہنا ہے۔ چنانچہ مقتدرہ نے اس معیاری آلے یا فرہنگ کو Electronic Dictionary of Localization of Computer Applications کے نام سے شائع کر دیا ہے۔ معیار بندی اور یکسانیت کے لیے یہ فرہنگ بنیادی حیثیت رکھتی ہے۔ انھی بنیادوں پر مائیکروسافٹ نے مجھ سے Style Guide، معیاری آلے کے طور پر Glossary اور Polycheck جیسے معیارات کی تکمیل کرائی جو وسٹا(Vista) اور دیگر سافٹ ویئر کے تراجم کے لیے بنیاد ٹھہرے۔

(۲۸۔ اگست ۲۰۰۵ء کو IBA کے عالمی سیمینار منعقدہ پرل کانٹی نینٹل ہوٹل کراچی اور ۲۹۔ اگست ۲۰۰۵ء کو مائیکروسافٹ کے پیش منظری اجلاس منعقدہ میریٹ ہوٹل کراچی میں پیش کیا گیا۔ "اخبارِ اُردو"، اسلام آباد میں اکتوبر ۲۰۰۵ء کو شائع ہوا اور جولائی ۲۰۰۸ء میں نظر ثانی کی گئی۔)

۷

اُردو میں ابلاغ اور جدید اطلاعیات

ہم اپنی بات کا آغاز اس خوشخبری سے کرتے ہیں کہ ہم نے کمپیوٹر سکرین کو اُردو میں بدل دیا ہے۔ مائیکروسافٹ ونڈوز ایکس پی اور سٹارٹر ایڈیشن نیز آفس ۲۰۰۳ء اب اُردو میں ہیں اور کمپیوٹر سیکھنے کے لیے انگریزی جاننا، ای میل یا ویب سائٹ قائم کرنے کے لیے حتیٰ کہ ویب سائٹ کا پتا (www) لکھنے کے لیے بھی اب رومن حروف درکار نہیں۔ اسی طرح موبائل فون نوکیا، موٹورولا بھی اب اُردو میں دستیاب ہیں۔ وسٹا (Vista) اُردو میں آرہی ہے۔ جدید اطلاعیات (Informatics) کے حوالے سے ابلاغیات میں یہ بہت بڑا انقلاب ہے اور ہم یہ انقلاب لا چکے ہیں۔

دنیا کی چار ہزار زبانوں میں سے صرف ۲۶ زبانیں ابھی تک کمپیوٹر کے ذریعے انٹرنیٹ پر آسکی ہیں اور صرف ۳۳ زبانوں کے مشینی ترجمہ کے سافٹ ویئر بن سکے ہیں۔ جبکہ دنیا میں ۶۶ زبانیں دفتری اور تعلیمی ضروریات پوری کر رہی ہیں۔ رسمِ خط کے لحاظ سے ابھی تک ساٹھ رسومِ خط تسلیم ہو سکے ہیں۔ دنیا کی زبانوں کو جدید ٹیکنالوجی کا دیو ہڑپ کیے جا رہا ہے۔ کہا جاتا ہے کہ اب صرف وہ زبان باقی بچے گی جو جدید اطلاعیات کی ٹیکنالوجی میں استعمال ہوگی۔ باقی زبانیں علمی دنیا سے حرفِ غلط کی طرح مٹ جائیں گی۔ یہ بات نوشتہ دیوار کی طرح اٹل ہے۔ اُردو کے حوالے سے تو اب یہ بھی کہا جا رہا ہے کہ اُردو اطلاعیات ہی اُردو کا مستقبل ہے۔

دنیا بھر میں اس شاہراہِ اطلاعیات نے دانش، تفکر، علم اور ابلاغ کے راستے ہی بدل ڈالے ہیں۔ کمپیوٹر اب تدریس کا بھی اور تحریر وصوت کا بھی سب سے بڑا واسطہ (Medium) بن چکا ہے اور بہت جلد شاعر اور نثر نگار کی جگہ بھی لینے والا ہے۔ کبوتر کی طرح محض آنکھیں بند کر لینے سے اس کا جادو سر چڑھ کر بولنے سے نہیں رہ سکتا۔ یہ باتیں ایک ان دیکھی طلسماتی دنیا کی لگتی ہیں۔ یوں لگتا ہے کہ ہزار سال پہلے کی ہزار داستان کا الف لیلوی ماحول اپنا حقیقی آغاز کر چکا ہے۔

اس بحث سے قطع نظر کہ ہماری زبان، ادب اور ثقافت کو اپنے ابلاغ کے لیے اُردو کے اس نئے واسطے کی ضرورت ہے یا نہیں اور آیا یہ آزاد فضا میں غیر ترقیاتی انداز میں زندہ رہ سکتی ہے یا نہیں، ہمیں اُردو کو اس جدید ٹیکنالوجی کے حوالے سے دیکھنے کی کوشش کرنا ہوگی۔

اُردو اطلاعیات کا شعبہ ۱۹۹۸ء سے کام کر رہا ہے اور اس میں کام کرتے ہوئے میں اگلے پچاس برس بعد کی اُردو کو اپنی آنکھوں سے دیکھ رہا، کانوں سے سُن رہا اور اگر کہیں تو زبان سے بول رہا ہوں۔ **"پاکستانی اُردو کے خدوخال"** میں جس "اُردش" قسم کی زبان کی پیش گوئی کی تھی، وہ اکیسویں صدی کے آغاز ہی میں اک کڑی دھوپ بن کر ہمارے سروں پر آ گئی ہے۔ جس ادبی محضر (Literary Discourse) میں ہم نے تربیتِ مطالعہ وتحریر پائی تھی، نئی نسلوں میں اس کا قاری عنقا ہو گیا ہے اور اب وہ نیا قاری جس زبان کا عادی ہو چکا ہے، ہمارا شاعر اور ادیب اس میں اظہار کی راہ نہیں پا رہا۔ قلم سے کاغذ پر لکھنے والے کلیدی تختے، برقیاتی گتے اور صوت نگار کی سان پر نہیں چڑھ رہے۔ ہم جس ادبی محضر اور علمی کینڈے (Academic Genre) کی بات بیسویں صدی میں کرتے تھے، وہ اک قصہء پارینہ بن چکا ہے۔ حرف اور عبارت کا جو واسطہ ہمارے زیرِ استعمال رہا ہے، اگلی نصف صدی میں شاید اس کی ضرورت ہی ختم ہو جائے۔ متن تا صوت اور صوت تا متن کی تمام ترمنتقلی کا کام کمپیوٹر سنبھال لے گا۔ ایک اندازے کے مطابق ہر ڈیڑھ سال بعد کمپیوٹر کا سائز دوگنا چھوٹا اور قوت دوگنا زیادہ ہو جاتی ہے۔ ایک چوتھائی صدی پہلے کے بڑے بڑے الماری نما کمپیوٹر بیسویں صدی کے آخر میں صرف میز پر اور اکیسویں صدی کے آغاز پر ہتھیلی میں آ چکے ہیں۔ اب اگلے عشرے تک یہ کان اور آنکھ پر اور اس کے بعد کے دو عشروں بعد صرف دماغ کے اندر نصب ہو جائیں گے۔ تب شاید انتقال وابلاغ کے لیے بول چال کی ضرورت بھی باقی نہیں رہے گی۔

ایسے میں اُردو کی صورت کیا ہو، اس کے ادبی وعلمی پہلوؤں کی تدریس کا اونٹ کس کروٹ بیٹھے گا، اس کا اندازہ کیا جا سکتا ہے۔

جدید اطلاعیات کا یہ الف لیلوی جنّ بوتل سے باہر آ چکا ہے مگر اس سے ڈرنے کی کوئی ضرورت نہیں۔ کوئی الہ دین ضرور ہوگا جو اس جنّ کو غلامی پر مجبور کر دے گا۔ یہ کام تو ہم سب کے کرنے کا ہے لیکن اگر سب ایسا نہیں کر سکتے تو کم از کم اخلاقی سطح پر اس الہ دین کو تقویت ہی بہم پہنچاتے رہیں۔ یہی اُردو کے حق میں بہتر ہے۔ وگرنہ اُردو کی داستاں تک بھی ۔۔۔۔۔

جدید ٹیکنالوجی کا یہ پہلو اُردو دوستوں سے دردمندی سے غور وفکر کا متقاضی ہے۔ اب تک ہم نے جن امور کو سطحی اور نچلے درجے کا سمجھا تھا اور آزاد فنون کی جن سطحوں کو بورژوائی مقام دے رکھا تھا، انھیں اتھل پتھل کرنے کی ضرورت ہے۔ سندھی لینگویج اتھارٹی اور پشتو اکیڈمی کو بھی میرا یہ مشورہ پہنچا تھا مگر ادبی نقار خانے اس نحیف طوطی کی آواز وہ شاید سن نہیں پا رہے۔

اُردو پر نوری نستعلیق کے حوالے سے ۱۹۸۰ء میں احمد مرزا جمیل کا ایک احسان ہوا۔ پھر بھارت میں "ان پیج" کے لفظ کار سافٹ ویئر سے ہماری کتابت کی بہت سی ضرورتیں پوری ہونے لگیں تو ہماری اکثریت مطمئن اور خوش ہو کر بیٹھ گئی کہ اُردو کمپیوٹر پر آ گئی ہے جبکہ وہ محض سکرین پر نظر آ رہی تھی۔ ہماری کتابیں اور

اخبارات اس پر دھڑا دھڑ چھپنے لگے۔ یہ حقیقت بہت کم لوگوں کو معلوم تھی کہ یہ سافٹ ویئر صرف لفظ کار ہیں، کمپیوٹر کار نہیں اور یہ صرف تصویری اساس پر عمل کرتے ہیں۔ یہ کمپیوٹر پر جگہ یا میموری ہزار گنا زیادہ گھیرتے ہیں۔ان کی رفتار انگریزی کے سافٹ ویئروں سے سو گنا کم ہے اور انٹرنیٹ/ ویب سائٹ/ ای میل وغیرہ کی دیگر ضرورتیں پوری نہیں کر پاتے۔ چنانچہ جب ہم نے کمپیوٹر سکرین پر اُردو کی جلد چڑھا دی تو مائیکروسافٹ نے اسے پیش کرنے کے لیے نسخ (تہامہ) فانٹ کا سہارا یہ کہہ کر لیا کہ ابھی نستعلیق کا کوئی موزوں کمپیوٹری فانٹ موجود نہیں۔ چنانچہ اک دوڑ لگ گئی۔اس سے پہلے نادرا کو شناختی کارڈ بنانے کے لیے اُردو کوائف کی تلاش کے لیے اُردو کے فانٹ کی ضرورت تھی۔ترتیب، تدوین اور پیش کش کے کئی مسائل درپیش تھے۔ کچھ طالع آزما کمپیوٹر سائنس کی دنیا سے اٹھے اور اُردو اطلاعیات کی اس لیڈرشپ پر قبضہ کرلیا۔ معاملات پھر بھی جوں کے توں رہے۔ضرورت اُردو کے ایسے ماہرِ لسان کی تھی جو ٹیکنالوجی کے ادق مسائل سے بھی واقف ہو اور نچلی سطح کے ان امور پر اعلیٰ تفکر و تدبر بھی کر سکے۔ اُردو کے ادیبوں اور پروفیسروں سے رجوع کیا گیا مگر وہاں تو انتقادی سطور پیش پیش تھیں اور صرف تفہیم و تشریح کی بات تھی۔ تجزیے، تقابل اور ڈیزائن تو کوسوں دور تھے۔ ایسا کام تو مستریوں اور کمپوڈروں کے کرنے کا تھا۔

مقتدرہ قومی زبان کا مرکز فضیلت برائے اُردو اطلاعیات اس میدان میں آگے بڑھا۔اس کے عزائم تو بہت بلند تھے۔ فی الوقت اس کی کارکردگی یہ ہے کہ اس نے مائیکروسافٹ جیسے اداروں کے ساتھ مل کر بہت سے معیاری اوزار وضع کیے ہیں، جن سے کمپیوٹر سکرین اُردو میں بدل لی ہے۔ اُردو میں ای میل ممکن ہوئی ہے۔ اُردو ویب سائٹ، برقیاتی کتاب، برقیاتی اشاعت اور برقیاتی رابطوں کی صورتیں سامنے آئی ہیں۔ ایک بنیادی اُردو فانٹ "پاک نستعلیق" کے نام سے ایجاد کیا گیا ہے، جو خالی حروف (Ghost Character) کی بنیاد پر تمام پاکستانی زبانوں کو ایک ہی وقت میں پیش کرنے کی صلاحیت رکھتا ہے۔ یہ ان پیج سے نوے گنا تیز رفتار ہے اور ہزار گنا کم جگہ گھیرتا ہے۔اس میں خالی جگہ حروف + نقاط اور شوشے استعمال کیے گئے ہیں۔اب کوئی پاکستانی یا عربی حروف میں لکھی جانے والی زبان اس کی دسترس سے باہر نہیں۔ اس مرکز کی طرف سے اب مشینی ترجمے کا سافٹ ویئر پیش کیا جانے والا ہے۔ پہلا مرحلہ دفتری انگریزی اُردو کا ہے۔ اگلا مرحلہ سائنسی و تکنیکی، پھر صحافتی اور شاید آخری مرحلہ ادبی ترجمے کا ہوگا۔ لطف کی بات یہ ہے کہ تکنیکی ترجمے کا سافٹ ویئر بنانا آسان ہے اور ادبی ترجمہ سب سے مشکل کام ہے۔

ادیبوں کے لیے کمپیوٹر پر کام کرنے میں حائل کلیدی تختہ کو بہت جلد ختم کر کے "تحریری گتہ" بنا کر دیا جارہا ہے، جو قلم کاری کو کمپیوٹر کی لفظ کاری میں بدل دے گا اور یوں ادیب کاغذ کی بجائے براہ راست بلا کسی کلیدی تختے کے نوری قلم کے ذریعے سکرین پر لکھ سکیں گے اور اسے کسی بھی فانٹ میں تبدیل کر سکیں گے۔ تحریری دستاویزات عکسی (Scan) ہو کر لفظ کاری میں بدل سکیں گی۔ قدیم دستاویزات اور کتابیں آسانی سے

کمپیوٹر سے عمل کار ہوسکیں گی۔ اگلے مرحلوں میں ہاتھ کی بجائے آواز کے ذریعے تحریر اور تحریر سے بات چیت کا عمل بھی انجام دے لیا جائے گا۔

مرکز کا یہ پراجیکٹ بہت جلد انسٹی ٹیوٹ کا درجہ پانے والا ہے۔ جو اُردو اطلاعیات میں ماہرین تیار کرنے کے لیے مابعد ڈاکٹریٹ تک کی تدریس و تحقیق بھی انجام دے گا۔

اُردو کا ایک بہت بڑا ذخیرہ (Bank) وضع کیا جا رہا ہے جو ایک کوائفیہ (Database) کی صورت میں اُردو کے کسی لفظ، ترکیب، محاورے، جملے، مصرعے، شعر کو اس کے معنی، مفہوم، ربط، تعلق، تعدد، حوالے کے ساتھ مابعد لغات (Post-lexico graphic) امور کے طور پر پیش کر سکے گا۔ یہ عمل برعکس بھی ہو گا۔ یعنی خیال سے معنی کا سفر بھی۔ ایک اندازے کے مطابق اس میں ساڑھے تین لاکھ الفاظ و اصطلاحات، چھ کروڑ اشعار اور دو کروڑ سے زائد جملے اور فقرے ہوں گے۔ اُردو کا یہ کوائفیہ وضع کر لیا گیا ہے۔ اب صرف مطلوبہ معلومات/کوائف کا اندراج باقی ہے اور یہ سب کچھ مفت میسر ہے اور رہو گا۔

اُردو کے لیے جدید ٹیکنالوجی کے تحقیقی و تدریسی امور کا احاطہ مندرجہ ذیل پہلوؤں سے کیا جانا چاہیئے:۔

۱۔رسم خط	۲۔متن	۳۔ترجمہ	۴۔صوتی رابطے
۵۔ذخائر	۶۔اشاعت	۷۔تدریس	۸۔تحقیق

رسمِ خط میں بہت سے امور شامل ہیں۔ جن کی تعریف، تصریح اور تحقیق درکار ہے۔ مرکز فضیلت اُردو کے حوالے سے نستعلیق رسمِ خط پر کام کر رہا ہے۔ اس کے دائرۂ کار میں حروف تہجی (صرف مصمتوں کی اکائیاں)، ابجد (مصمتوں اور مصوتوں کی بنیادی اکائیاں)، ہجا (ایک یا زائد مصمتے اور ایک مصوتے کا جوڑ)، ابوجدا (حروف و ہجا کا ملا جلا سینٹ)، ترسیمہ (مل کر لکھے گئے حروف جو ا، ر، د، و، پر ختم ہوں)، ترقیمہ (نشانات)، علامتیات، رموز اوقاف، ترتیبِ حروف، جوڑ کار (Joinder)، سمتِ تحریر، معنویاتِ حروف، اندازِ عبارت، فاصلِ تحریر، دندانے، شوشے، حرکات، صوتیاتی الفبائی حروف، ہندسے، فاصلِ تاریخ، کرنسی علامات، اضافی علامات، آرائش، فاصلہ جات، رجوعی نقلِ حرفی اور دیگر کئی متفرق امور شامل ہیں۔ جنھیں موضوعی کی بجائے معروضی اور رجوعی انداز میں معیار بند کرنا ضروری ہے۔ اس مقصد کے لیے عالمی معیارات وضع کیے جا رہے ہیں۔ حروفِ ابجد کی معیار بندی کی گئی ہے۔ کئی "رہنمائے وضع" بنائی گئی ہیں۔ عالمی ضابطوں کے لیے یونی کوڈ تک کی رکنیت حاصل کی گئی اور اس میں اُردو کے لیے تبدیلی کی کوششیں کی گئی ہیں۔

دوسرا مرحلہ ادخالِ متن کا ہے، جس سے الفاظ اور ان کی حد، بنیادی ذخیرہء الفاظ، مجازی ذخیرہء الفاظ، معجمی (تھیسارس کا) ذخیرہء الفاظ، معنویاتی روابط، معیاری املا، معیاری ہجے، صوتیے، مصوتے، مصمتے، سادہ جملے کی حدود، افعال اور زمانے اور ان میں استثنا، مرکب جملے کی جدولیں، حدودِ جملہ و حدودِ فقرہ، نحو، طے

شدہ (default) نحو، استثنا، متبادل نحو، کمپیوٹری لسانیات، کمپیوٹری ترتیبات، طرزیات (Syslistics)، تدوینِ متن، قواعد کی جنس، تعداد، املا کے لحاظ سے لفظ کی صورتیں، تصوریات (Conceptology)، معجمی یا تھیسارس کے حوالے سے معنوی تعلق، تلازماتِ لفظی (Word Associations)، معنویات کی سطحیں، بول چال کے سانچے (Registers)، کینڈے (Genres)، متنیت (Textuality)، توسل (Mediation)، زبانی وتحریری محضری اور ثقافتی انداز اور تجزیے، وضعی، لغوی اور مستعمل صورتیں، کمپیو نظم کے سانچے (صدر، ضرب، ردیف، وزن، بحور)، کمپیوانشا (۳ تا ۵ پیرا موضوعات، تصوریات)، حدودِ تحریر وغیرہ کئی پہلو شامل ہیں۔ ان کے لیے متعدد فہرستیں اپنی خصوصیات کے ساتھ مرتب ہونی چاہئیں۔

تیسرا مرحلہ ترجمہ کی ایسی مشین بنانے کا ہے جو خودکار ترجمے اور ترجمانی کے کام آ سکے۔ اس میں نقل یا ترسیمِ حرفی (Transliterational Graphics) سے لے کر لفظ کی خاصیت، علمِ ترجمہ، سادہ جملوں کا تقابلی مقام، مرکب جملوں کی تقسیم اور انگریزی سے تقابل اور منتقلی کے فارمولے، استثنا، ترتیبِ نو اور اس کے اضافی امور، مثال اساس (Example Based) ترجمے اور علم اساس (Knowledge Based) ترجمے کے تقاضے، محضری مشینی ترجمہ، کمپیوٹرا عانتی ترجمہ (CAT)، خودکار ترجمہ مشین (MT) اور ان کے لیے درکار کمپیوٹرا اساس لغات اور فرہنگیں بنانا مقصود ہیں۔ اگرچہ کئی معیاری فرہنگیں اور جدید جدولیں تیار کی گئی ہیں۔ دنیا میں ترجمے کے کئی سافٹ ویئر (TRADOS وغیرہ) موجود ہیں مگر ابھی ایک ہفت خواں طے کرنا باقی ہے۔

چوتھے مرحلے میں آواز سے حروف اور حروف والفاظ سے آواز تک تبدیلی کے لیے تعبیر وخاصیت کے حوالے سے معیاری فرہنگوں، صوتی تجزیوں، صوتی ابجدوں، لہجوں اور شدتوں کی پیمائش کی ضرورت ہے۔ اس سے صوت نگار (Phonogram) وجود میں آئیں گے۔ فاسٹ یونیورسٹی لاہور کے طلبہ نے ایسے کئی منصوبے انجام دیے ہیں۔ "اخبارِ اُردو" میں ان کی اشاعت ہوتی رہی ہے۔

پانچویں شے اُردو کوائفیہ ذخائر کے حوالے سے بنیادی اہمیت رکھتا ہے۔ اُردو کے کوائفیہ کو آگے چل کے کثیر لسانی کوائفیوں میں بدلنا، پانچ لاکھ پچھتر ہزار اصطلاحات اور ایک کروڑ ستر لاکھ تصورات کی اُردو میں منتقلی اور باہمی ربط کے ساتھ ذخیرہ کاری ایک بہت بڑا چیلنج ہے۔ یوروپی کمیشن کا ایک ایسا مرکز نوزبانوں کے لیے لکسمبرگ میں کام کر رہا ہے اس کے تعاون سے اُردو کوائفیہ بھی یہ تمام تصورات پیش کر سکے گا۔

کمپیوٹر کے ذریعے اشاعتی امور اتنے سادہ نہیں جتنے اُردو دنیا کو نظر آتے رہے ہیں۔ اشاعت کی (DTP) ضروریات ان پیج وغیرہ جیسے سافٹ ویئر سے پوری نہیں ہو پاتیں۔ اس پہلو پر ایڈوب (Adobe) کی مداریات (Acrobatics) کی ضرورت ہے۔ جو پی ڈی ایف (PDF) اور دیگر لفظ

کاری مسلیں اور پوشے (Word Processing Files and Folders) حسبِ ضرورت وضع کر سکے۔ مرکز فضیلت بہت جلد اس میں بھی پیش رفت حاصل کر لے گا۔

ساتویں بات یہ کہ اُردو میں اعلیٰ سطح پر تدریس کے حوالے سے کئی طرح کے سافٹ ویئر وجود میں لانے کے لیے اُردو کی بنیادی تحقیق درکار ہے جس میں معیاری متن کی تشکیل، تدریسی اسباق، متعامل (Interactive) اسباق، مشینی کہانی، لفظی بنت کار (Weaver)، کمپیونظم، کمپیوانشا، کمپیومکالمہ، تدریسِ خواندگی، اختلافِ انتقاد، امکاناتِ تدریس، تدریسی کینڈے وغیرہ کے متعدد تشریحی پہلو شامل ہیں۔ یہ کام اُردو دانوں، اُردو کے شعبوں اور اکادمیوں کے کرنے کا ہے۔ اس سے ابلاغیات، ادبیات اور تعلیم کی تکنیکی پہلوؤں سے ترقی اور مہارت کے کئی گوشے وا ہوں گے۔

اُردو کے دوستوں نے تو ابھی تحقیق کی وادی میں قدم نہیں رکھا تھا اور اب ہائر ایجوکیشن کمیشن کے ہاتھوں پریشان ہو رہے ہیں۔ تحقیق ایک طریقِ کار ہے اور محض اسی طریق کار پر عمل پیرا ہو کر ہی تحقیق انجام دی جا سکتی ہے۔ اُردو کے ساتھ ساتھ اُردو اطلاعیات کے مندرجہ بالا سمیت بہت سے امور میں تحقیق درکار ہے۔ اصولِ تحقیق کے کئی انگریزی سافٹ ویئر (SPSS وغیرہ) موجود ہیں جو کوائف کے باہمی تعلق کی جدولیں وضع کرتے ہیں، انھیں اُردو میں ڈھالا جانا مقصود ہے۔ کمپیوٹری لسانیات پر جو تھوڑا بہت کام دیگر اداروں میں ہو رہا ہے، اس میں پشاور یونیورسٹی، فاسٹ لاہور اور اسلامیہ یونیورسٹی اسلام آباد کے کمپیوٹر سائنس کے شعبوں میں قدرے تحقیق انجام پا رہی ہے۔

اب اگر ایک نظر ان ضمیموں پر ڈالی جائے جو اس مقالے کے ساتھ منسلک ہیں تو ذرا زیادہ وضاحت ہو سکے گی۔ پہلے ضمیمے میں اُردو کے معیاری حروفِ تہجی اور اس کی ترتیب علّت دی گئی ہے، جو مقتدرہ قومی زبان کی ویب سائٹ www.nla.gov.pk پر بھی موجود ہے۔ اس کی معیار بندی پر پانچ برس صرف ہوئے۔ اب یہ رائج ہو چکے ہیں۔

دوسرے ضمیمے پر اُردو کلیدی تختہ ہے جو مقتدرہ قومی زبان کی طرف سے معیار بند کیا گیا ہے۔ یہ کلیدی تختہ نادرا کے زیرِ استعمال ہے اور جزوی طور پر مائیکروسافٹ کے انگریزی کے ونڈوز ایکس پی اور اُردو کے ونڈوز ایکس پی میں شامل ہے۔

تیسرے ضمیمے میں یونی کوڈ پر موجود عربی حروف کا صفحہ ورژن نمبر ۳ دیا گیا ہے۔ جس میں مقتدرہ کی طرف سے تبدیلی ورژن نمبر ۴ میں واضح ہے۔ اس تبدیلی میں علاماتِ صفر، شاعر، حوالہ، درود و تقدیس اور ادبیات کی علامات اور خالی (Ghost) کشتیاں یا حروف شامل ہیں۔ جن سے نقاط کے بغیر قرآن اور دیگر

عبارات کی لفظ کاری ممکن ہوگئی ہے۔ یونی کوڈ ورژن ۵ بھی ملاحظہ ہوسکتا ہے۔ جس میں ۰۶ پر جگہ ختم ہونے کے باعث حروف ۰۷ پر دیے گئے ہیں مگر جگہ پھر بھی ختم ہوگئی۔ اب جگہ کا مسئلہ پیدا ہوگیا۔ اس کا حل بنیادی حرف اور خالی نقاط میں پوشیدہ ہے۔ خالی نقاط کی بحث یونی کوڈ کے ساتھ جاری رہی جو مرکز فضیلت کی ویب سائٹ www.nlauit.gov.pk پر دیکھی جا سکتی ہے۔ ۱۱ مئی ۲۰۰۸ء کو UTC نے اسے قبول کرلیا اور ۰۸ پر جگہ دے دی۔ یہ یونی کوڈ ۶ میں آئے گی۔

چوتھے ضمیمے میں یونی کوڈ پر موجود ساٹھ رسومِ خط کے نام دیے گئے ہیں۔ ان میں سے پچاس کو باقاعدہ ضابطہ صفحات حاصل ہوئے ہیں اور ان میں سے ۴۳ جدید رسومِ خط ہیں۔ نستعلیق ان میں کہیں شامل نہیں کیونکہ اسے غلطی سے نسخ کا ذیلی سیٹ سمجھا جاتا ہے جبکہ اس کا ترسیمہ ح، م، ی کے حوالے نیز نشست اور مقام کے حوالے سے مختلف ہوتا ہے۔ اُردو کو ویسے بھی عربی کا ذیلی سیٹ ہی سمجھا جاتا ہے، جو غلط ہے۔

پانچویں ضمیمے میں سادہ جملوں کا فارمولا دیا گیا ہے جو چومسکی کے نقطۂ نظر سے وضع ہوتا ہے۔

چھٹے ضمیمے میں ترجمے کے لیے مرکب جملوں کا بنیادی فارمولا، اس کی ترتیبِ نو کے مسائل کے ساتھ دیا گیا ہے۔

ساتویں ضمیمے میں اُردو کوائفیہ میں معلومات داخل کرنے کا فارم شامل ہے۔

آخری تینوں ضمیموں سے کمپیوٹر کے لیے اُردو قواعد کی ورکشاپ اسلام آباد ۹ تا ۱۱ مارچ ۲۰۰۷ء میں بنیادی کام لیا گیا تھا اور انھیں حتمی شکل دی گئی ہے۔ جس کی بنیاد پر مشینی ترجمے اور اُردو کوائف کے سافٹ ویئر تیار کیے جارہے ہیں۔

اُردو کو اس کے ابلاغی پہلوؤں سے تکنیکی ترقی کے تیز ترعمل میں شامل کرنے کے لیے ضروری ہے کہ ادب، صحافت اور میڈیا سے متعلقہ لوگ ان گزارشات پر مفصل غور کریں جنھیں بے حد اختصار کے ساتھ یہاں پیش کیا گیا ہے۔

یہ کام ہورہے ہیں یا ہوجائیں گے لیکن ضرورت اس امر کی ہے کہ اُردو کے سلسلے میں تدریس اور نصابات کا رخ موڑا جائے۔ ہماری تحریک پر ۲۰۰۳ء سے ہائر ایجوکیشن کمیشن نے اُردو کے اعلیٰ سطحی نصاب میں تبدیلیاں کی ہیں اور اُردو اطلاعیات کو اس کا حصہ بنایا، مگر ابھی تک کسی یونیورسٹی کے شعبہ اُردو نے اس پر عمل شروع نہیں کیا۔ اُردو لسانیات کی تدریس کا بھی کہیں آغاز نہیں ہوا۔ پھر اُردو کے ماہرینِ کمپیوٹری لسانیات کہاں سے آئیں گے؟ اُردو کے مختلف متون اور ادبی محضرات کے تجزیے کہاں چھپیں گے۔ شعبہ ہائے اُردو تو ماضی پرستی میں مست اپنی "بقیہ زندگی" بسر کرنے میں مصروف ہیں۔ حال کی اُردو اور مستقبل کی ضرورتوں پر کون

تحقیق کرے گا۔ یہ تحقیق سائنسی بنیادوں پر کرنے کی ضرورت ہے۔ تحقیقی منصوبوں کے لیے رقم کی کوئی کمی نہیں مگر یہ کام کہاں انجام پائے گا اور یہ جو کام ہو رہے ہیں ان کی تدریس کہاں ہو پائے گی۔ ایک ادارہ کافی نہیں۔ اُردو اطلاعیات کا شعبہ قائم کرنے کی کوشش کی جارہی ہے مگر بادِ مخالف کی سم آلودگی؟

(اُردو کانفرنس، لاہور میں ۵۔ مئی ۲۰۰۷ء کو پڑھا گیا۔ "اخبارِ اُردو" اسلام آباد، مئی ۲۰۰۷ء اور "مخزن" لاہور شمارہ ۱۴، مئی ۲۰۰۷ء میں شائع ہوا۔ جولائی ۲۰۰۸ء میں نظرِ ثانی کی گئی۔)

ضمیمہ:۱

اُردو کے حروفِ تہجی اور ان کی ترتیب

ا آ ب بھ پ پھ ت تھ ٹ ٹھ

ث ج جھ چ چھ ح خ د دھ ڈ ڈھ

ذ ر رھ ڑ ڑھ ز ژ س ش ص ض ط ظ ع غ

ف ق ک کھ گ گھ ل لھ م مھ ن نھ ں

ںٛھ و وھ

ہ ۃ ء ی یھ ے

ترتیب علامات: ا اَ اِ اُ اٗ اٖ اً اٍ اٌ اَ اِ آ

اُردو اور پاکستانی زبانوں کے ۲۲ بنیادی حروفِ تحریر (بلا نقاط اور شوشے)

ا ب ح د ر س ط ص ع ف ق ک گ ل م ں و ہ ء ی ھ ے

ضمیمہ ۲:

مقتدرہ کا کلیدی تختہ

مائیکروسافٹ ونڈوز XP میں شامل اُردو کلیدی تختہ

ضمیمہ:۳

یونی کوڈ ورژن ۔۳

Arabic

	060	061	062	063	064	065
0				ذ 0630	ـ 0640	◌ِ 0650
1			ء 0621	ر 0631	ف 0641	◌ّ 0651
2			آ 0622	ز 0632	ق 0642	◌ْ 0652
3			أ 0623	س 0633	ك 0643	◌ٓ 0653
4			ؤ 0624	ش 0634	ل 0644	◌ٔ 0654
5			إ 0625	ص 0635	م 0645	◌ٕ 0655
6			ئ 0626	ض 0636	ن 0646	
7			ا 0627	ط 0637	ه 0647	
8			ب 0628	ظ 0638	و 0648	
9			ة 0629	ع 0639	ى 0649	
A			ت 062A	غ 063A	ي 064A	
B		؛ 061B	ث 062B		◌ً 064B	
C	، 060C		ج 062C		◌ٌ 064C	
D			ح 062D		◌ٍ 064D	
E			خ 062E		◌َ 064E	
F		؟ 061F	د 062F		◌ُ 064F	

Arabic

	068	069	06A	06B	06C	06D	06E	06F
0	ڀ 0680	ڐ 0690	ڠ 06A0	ڰ 06B0	ۀ 06C0	ې 06D0	◌۠ 06E0	۰ 06F0
1	ځ 0681	ڑ 0691	ڡ 06A1	ڱ 06B1	ہ 06C1	ۑ 06D1	◌ۡ 06E1	۱ 06F1
2	ڂ 0682	ڒ 0692	ڢ 06A2	ڲ 06B2	ۂ 06C2	ے 06D2	◌ۢ 06E2	۲ 06F2
3	ڃ 0683	ړ 0693	ڣ 06A3	ڳ 06B3	ۃ 06C3	ۓ 06D3	◌ۣ 06E3	۳ 06F3
4	ڄ 0684	ڔ 0694	ڤ 06A4	ڴ 06B4	ۄ 06C4	۔ 06D4	◌ۤ 06E4	۴ 06F4
5	څ 0685	ڕ 0695	ڥ 06A5	ڵ 06B5	ۅ 06C5	ە 06D5	ۥ 06E5	۵ 06F5
6	چ 0686	ږ 0696	ڦ 06A6	ڶ 06B6	ۆ 06C6	◌ۖ 06D6	ۦ 06E6	۶ 06F6
7	ڇ 0687	ڗ 0697	ڧ 06A7	ڷ 06B7	ۇ 06C7	◌ۗ 06D7	◌ۧ 06E7	۷ 06F7
8	ڈ 0688	ژ 0698	ڨ 06A8	ڸ 06B8	ۈ 06C8	◌ۘ 06D8	◌ۨ 06E8	۸ 06F8
9	ډ 0689	ڙ 0699	ک 06A9	ڹ 06B9	ۉ 06C9	◌ۙ 06D9	۩ 06E9	۹ 06F9
A	ڊ 068A	ښ 069A	ڪ 06AA	ں 06BA	ۊ 06CA	◌ۚ 06DA	◌۪ 06EA	ۺ 06FA
B	ڋ 068B	ڛ 069B	ګ 06AB	ڻ 06BB	ۋ 06CB	◌ۛ 06DB	◌۫ 06EB	ۻ 06FB
C	ڌ 068C	ڜ 069C	ڬ 06AC	ڼ 06BC	ی 06CC	◌ۜ 06DC	◌۬ 06EC	ۼ 06FC
D	ڍ 068D	ڝ 069D	ڭ 06AD	ڽ 06BD	ۍ 06CD	۝ 06DD	◌ۭ 06ED	۽ 06FD
E	ڎ 068E	ڞ 069E	ڮ 06AE	ھ 06BE	ێ 06CE	۞ 06DE		۾ 06FE
F	ڏ 068F	ڟ 069F	گ 06AF	ڿ 06BF	ۏ 06CF	◌۟ 06DF		

یونی کوڈ ورژن۔۴

0600 **Arabic** **06FF**

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	؀ 0600	◌ؐ 0610		ذ 0630	ـ 0640	◌ِ 0650	٠ 0660	◌ٰ 0670	ڀ 0680	ڐ 0690	ڠ 06A0	ڰ 06B0	ۀ 06C0	ې 06D0	◌۠ 06E0	۰ 06F0
1	؁ 0601	◌ؑ 0611	ء 0621	ر 0631	ف 0641	◌ّ 0651	١ 0661	ٱ 0671	ځ 0681	ڑ 0691	ڡ 06A1	ڱ 06B1	ہ 06C1	ۑ 06D1	◌ۡ 06E1	۱ 06F1
2	؂ 0602	◌ؒ 0612	آ 0622	ز 0632	ق 0642	◌ْ 0652	٢ 0662	ٲ 0672	ڂ 0682	ڒ 0692	ڢ 06A2	ڲ 06B2	ۂ 06C2	ے 06D2	◌ۢ 06E2	۲ 06F2
3	؃ 0603	◌ؓ 0613	أ 0623	س 0633	ك 0643	◌ٓ 0653	٣ 0663	ٳ 0673	ڃ 0683	ړ 0693	ڣ 06A3	ڳ 06B3	ۃ 06C3	ۓ 06D3	◌ۣ 06E3	۳ 06F3
4		◌ؔ 0614	ؤ 0624	ش 0634	ل 0644	◌ٔ 0654	٤ 0664	ٴ 0674	ڄ 0684	ڔ 0694	ڤ 06A4	ڴ 06B4	ۄ 06C4	۔ 06D4	◌ۤ 06E4	۴ 06F4
5		◌ؕ 0615	إ 0625	ص 0635	م 0645	◌ٕ 0655	٥ 0665	ٵ 0675	څ 0685	ڕ 0695	ڥ 06A5	ڵ 06B5	ۅ 06C5	ە 06D5	ۥ 06E5	۵ 06F5
6			ئ 0626	ض 0636	ن 0646	◌ٖ 0656	٦ 0666	ٶ 0676	چ 0686	ږ 0696	ڦ 06A6	ڶ 06B6	ۆ 06C6	◌ۖ 06D6	ۦ 06E6	۶ 06F6
7			ا 0627	ط 0637	ه 0647	◌ٗ 0657	٧ 0667	ٷ 0677	ڇ 0687	ڗ 0697	ڧ 06A7	ڷ 06B7	ۇ 06C7	◌ۗ 06D7	◌ۧ 06E7	۷ 06F7
8			ب 0628	ظ 0638	و 0648	◌٘ 0658	٨ 0668	ٸ 0678	ڈ 0688	ژ 0698	ڨ 06A8	ڸ 06B8	ۈ 06C8	◌ۘ 06D8	◌ۨ 06E8	۸ 06F8
9			ة 0629	ع 0639	ى 0649		٩ 0669	ٹ 0679	ډ 0689	ڙ 0699	ک 06A9	ڹ 06B9	ۉ 06C9	◌ۙ 06D9	۩ 06E9	۹ 06F9
A			ت 062A	غ 063A	ي 064A		٪ 066A	ٺ 067A	ڊ 068A	ښ 069A	ڪ 06AA	ں 06BA	ۊ 06CA	◌ۚ 06DA	◌۪ 06EA	ۺ 06FA
B		؛ 061B	ث 062B		◌ً 064B		٫ 066B	ٻ 067B	ڋ 068B	ڛ 069B	ګ 06AB	ڻ 06BB	ۋ 06CB	◌ۛ 06DB	◌۫ 06EB	ۻ 06FB
C	، 060C		ج 062C		◌ٌ 064C		٬ 066C	ټ 067C	ڌ 068C	ڜ 069C	ڬ 06AC	ڼ 06BC	ی 06CC	◌ۜ 06DC	◌۬ 06EC	ۼ 06FC
D	؍ 060D		ح 062D		◌ٍ 064D		٭ 066D	ٽ 067D	ڍ 068D	ڝ 069D	ڭ 06AD	ڽ 06BD	ۍ 06CD	۝ 06DD	◌ۭ 06ED	۽ 06FD
E	؎ 060E		خ 062E		◌َ 064E		ٮ 066E	پ 067E	ڎ 068E	ڞ 069E	ڮ 06AE	ھ 06BE	ێ 06CE	۞ 06DE	ۮ 06EE	۾ 06FE
F	؏ 060F	؟ 061F	د 062F		◌ُ 064F		ٯ 066F	ٿ 067F	ڏ 068F	ڟ 069F	گ 06AF	ڿ 06BF	ۏ 06CF	◌۟ 06DF	ۯ 06EF	ۿ 06FF

0750 **Arabic Supplement** **077F**

	075	076	077
0	ݐ 0750	ݠ 0760	ݰ 0770
1	ݑ 0751	ݡ 0761	ݱ 0771
2	ݒ 0752	ݢ 0762	ݲ 0772
3	ݓ 0753	ݣ 0763	ݳ 0773
4	ݔ 0754	ݤ 0764	ݴ 0774
5	ݕ 0755	ݥ 0765	ݵ 0775
6	ݖ 0756	ݦ 0766	ݶ 0776
7	ݗ 0757	ݧ 0767	ݷ 0777
8	ݘ 0758	ݨ 0768	ݸ 0778
9	ݙ 0759	ݩ 0769	ݹ 0779
A	ݚ 075A	ݪ 076A	ݺ 077A
B	ݛ 075B	ݫ 076B	ݻ 077B
C	ݜ 075C	ݬ 076C	ݼ 077C
D	ݝ 075D	ݭ 076D	ݽ 077D
E	ݞ 075E	ݮ 076E	ݾ 077E
F	ݟ 075F	ݯ 076F	ݿ 077F

60

0880 | Arabic Glyph Parts | 089F

	x88	x89
0	xx80	xx90
1	xx81	xx91
2	xx82	xx92
3	xx83	xx93
4	xx84	xx94
5	xx85	xx95
6	xx86	
7	xx87	
8	xx88	
9	xx89	
A	xx8A	
B	xx8B	
C	xx8C	
D	xx8D	
E	xx8E	
F	xx8F	

0880 ARABIC SINGLE NUQTA ABOVE
0881 ARABIC SINGLE NUQTA BELOW
0882 ARABIC DOUBLE NUQTA ABOVE
0883 ARABIC DOUBLE NUQTA BELOW
0884 ARABIC TRIPLE NUQTA ABOVE
0885 ARABIC TRIPLE NUQTA BELOW
0886 ARABIC TRIPLE INVERTED NUQTA ABOVE
0887 ARABIC TRIPLE INVERTED NUQTA BELOW
0888 ARABIC QUADRUPLE NUQTA ABOVE
• Sindhi
0889 ARABIC QUADRUPLE NUQTA BELOW
• Sindhi
088A ARABIC DOUBLE DANDA ABOVE
• Sindhi
088B ARABIC DOUBLE DANDA BELOW
• Sindhi
088C ARABIC DOUBLE NUQTA VERTICAL ABOVE
• Sindhi
088D ARABIC DOUBLE NUQTA VERTICAL BELOW
• Sindhi
088E ARABIC SINGLE KASHIDA ABOVE
• Urdu
088F ARABIC SINGLE KASHIDA BELOW
• Urdu
0890 ARABIC DOUBLE KASHIDA ABOVE
• Urdu
0891 ARABIC DOUBLE KASHIDA BELOW
• Urdu
0892 ARABIC SINGLE CIRCLE ABOVE
•Pashto
0893 ARABIC SINGLE CIRCLE BELOW
•Pashto
0894 ARABIC TOTA ABOVE
• Urdu
0895 ARABIC TOTA BELOW
• Urdu

Printed using UniBook™
(http://www.unicode.org/unibook/)

Date: 17-Apr-2008

1

ضمیمہ:۴

یونی کوڈ کے ساٹھ رسوم خط

دنیا میں اس وقت تک کمپیوٹر کے حوالے سے یونیکوڈ میں صرف ساٹھ رسومِ خط تسلیم کیے گئے ہیں:

۱۔لاطینی	۲۔یونانی	۳۔قبطی	۴۔سیریلی	۵۔گیلگولیتھی
۶۔آرمینیائی	۷۔جارجیائی	۸۔یونانی	۹۔عربی	۱۰۔شامی
۱۱۔تھانا	۱۲۔دیوناگری	۱۳۔بنگالی	۱۴۔گورمکھی	۱۵۔گجراتی
۱۶۔اڑیا	۱۷۔تامل	۱۸۔تلگو	۱۹۔کنٹری	۲۰۔ملیالم
۲۱۔سنہالا	۲۲۔تبتّی	۲۳۔فاگس پا	۲۴۔لیمبو	۲۵۔سلوتی ناگری
۲۶۔خروشتی	۲۷۔تھائی	۲۸۔لاؤ	۲۹۔میانمار	۳۰۔خمیر
۳۱۔تائی لی	۳۲۔نئ تائی لیو	۳۳۔فلپائنی	۳۴۔بوگنی	۳۵۔بالینی
۳۶۔ہین(CJK)	۳۷۔بوپوموفو	۳۸۔ہیراگانا	۳۹۔کٹاکانا	۴۰۔ ہنگول
۴۱۔یی	۴۲۔ایتھوپیائی	۴۳۔ منگولیائی	۴۴۔عثمانیہ/صومالی	۴۵۔ٹھی ناغ
۴۶۔ این کو	۴۷۔شیروکی	۴۸۔کینیڈااقدیم	۴۹۔صحرائی	۵۰۔شاوین
۵۱۔اوغم	۵۲۔قدیم اطالوی	۵۳۔رونی	۵۴۔گوتھی	۵۵۔لائیز
۵۶۔کیپریوٹ	۵۷۔فونیقی	۵۸۔ اگاریتی	۵۹۔قدیم پارسی	۶۰۔سامری/ اکادی

ضمیمہ:۵

نوم چامسکی کے نظریات
Noam Chomsky

انتقالی قواعد:Transformational Rules

تجزیہ

مثلاً اگر W+X+A

W+Y+?+X+A = W+Y+A

جیسے: اس نے + روٹی + کھائی

اس نے + بوٹی + کھائی

= اس نے روٹی اور بوٹی کھائی

سوالیہ: "کیا" شروع میں اور نفی = نہیں فعل سے پہلے۔

ضمیمہ:٦

Recombination

ان سادہ فقروں کی ترتیب کے لیے مندرجہ ذیل دو فارمولے موجود ہیں۔

$$\frac{1}{N}a+n+(n-1)+(n-2)...+n+\frac{1}{V}b$$

بعض اوقات ایک مرکب جملہ دوسرے سے حروف عطف یا حروف جار کے ذریعے جڑا ہوتا ہے مثلاً or , and

اس کا فارمولا حسب ذیل ہے۔

$$\left(\frac{1}{N}a+n+(n-1)+(n-2)+...+\frac{1}{V}b\right)+[and/or]+\left(\frac{1}{N}a+n+(n-1)+(n-2)+...+\frac{1}{V}b\right)$$

ان دو فارمولوں کے علاوہ مزید کونسے فارمولے ہو سکتے ہیں؟ براہ کرم درج کریں۔

ضمیمہ:۷

اردو کوائفیہ فارم

اندراج ______________________

1	کلیدی الفاظ	
2	دیگر املا	
3	املا امالہ (اگر)	
4	تصریفی صورتیں	
5	ہم ردیف	

6	تجنیس حرفی نمبر	
6۔ الف	عمومی تلفظ	
7	ہم قافیہ	

<table>
<tr><td>8</td><td colspan="2">تجنیس صوتی نمبر</td><td colspan="3"></td></tr>
<tr><td rowspan="6">8- الف</td><td rowspan="6">الفاظ کی سطح / نوعیت</td><td></td><td></td><td>ہاں</td><td>نہیں</td></tr>
<tr><td>روزمرہ / بول چال</td><td></td><td></td><td></td></tr>
<tr><td>خواندہ</td><td></td><td></td><td></td></tr>
<tr><td>تعاملی</td><td></td><td></td><td></td></tr>
<tr><td>ادبی</td><td></td><td></td><td></td></tr>
<tr><td>اختصاصی</td><td></td><td></td><td></td></tr>
<tr><td>8۔ ب</td><td colspan="2">اشتقاق</td><td colspan="3"></td></tr>
<tr><td>8۔ ج</td><td colspan="2">قواعدی حیثیت</td><td colspan="3"></td></tr>
<tr><td>8۔ د</td><td colspan="2">جنس</td><td colspan="3"></td></tr>
</table>

8۔ ہ	متضاد جنس	
8۔ و	تکبیر	
8۔ ز	تصغیر	
8۔ ح	عدد	
8۔ ط	جمع / واحد	
8۔ ی	جمع الجمع	
9	مشتقات	

10	معنی نمبر		
10- الف	معنی	سلینگ / مقامی	
		عمومی	
		مجازی	
		اصطلاحی	
		سطح بالا	
		سطح زیریں	

11	متعلقات	
11- الف	روزمرے	
11- ب	محاورے	
11- ج	مصرعے	
11- د	اشعار	
11- ہ	ضرب الامثال	
11- و	تراکیب	

11- ز	مترادفات	
11 ـ ح	متضادات	
11- ط	متلازمات	
11- ع	اضافی معلومات	

12	مثال نمبر							
12- الف	مثال قسم	روزمرہ	محاورہ	جملہ	شعر	مصرع	ضرب المثل	دیگر
12- ب	مثال - متن							
12- ج	مصنف / شاعر							
12- د	کتاب / مضمون / رسالہ							
12- ہ	مرتب/مؤلف							
12- و	مقام اشاعت							
12- ز	ناشر							
12 ـ ح	سن تحریر/ اشاعت							
12- ط	جلد/ شمارہ							
12- ی	صفحہ نمبر							

اُردو میں کمپیوٹر کا تدریسی اطلاق

کمپیوٹر خاص طور پر لفظ کار (word processor) کے ذریعے اُردو کی تدریس شاید ایک قبل از وقت خیال ہو کہ اُردو تدریسات میں تو ابھی اُردو ٹائپ مشین بھی داخل نہیں ہوئی۔ لفظ کار کو تو صرف اُردو کتابت کا متبادل سمجھا جاتا ہے۔ اُردو میں کمپیوٹر کا استعمال بڑے پیمانے پر صرف اُردو شناختی کارڈوں کے حصول کے لیے سامنے آیا ہے۔ بینکوں، ڈاک خانوں، بلوں، ٹیلی فون، اور ریلوے کی ڈائریکٹریوں کو ابھی اُردو کی ہوا نہیں لگی۔ اُردو ادیب ابھی اُردو کلیدی تختے پر خود بھی نہیں بیٹھا۔ ایسے میں تدریس اُردو میں کمپیوٹر کا استعمال؟ ایک اچنبھے کی بات ہے۔ ایک دیوانے کا خواب ہے۔ آگے چل کر جب اُردو میں برقیاتی کتاب (E.Book) وجود میں آئے گی۔؟ تو خیال وخواب کی باتیں؟۔۔۔۔۔۔۔۔۔

تاہم جہاں کہیں، کسی اچھے سکول میں اُردو کی تدریس میں ذاتی کمپیوٹر یا جماعتی کمپیوٹر استعمال میں آئیں گے فی الوقت وہاں اُردو لفظ کار کا استعمال تو سکھایا جا سکتا ہے۔ لکھنے پڑھنے یعنی تصنیف و تالیف میں تو یہ لفظ کار مصنفوں کی بے حد مدد کر سکتا ہے کہ اگر ان کا لکھا ہوا مواد محفوظ ہو تو وہ اسے بے شمار صورتوں میں بدل سکتے ہیں۔ اگر کبھی اُردو کے کلاسیکی ادب کی سی ڈی بازار میں ملنے لگی تو اس کے اقتباسات نقل کرنے کی بجائے متن سے حاصل کر کے اپنے مسودے میں داخل کیے جا سکتے ہیں۔ اگر کبھی اُردو کی تدریسی کتابیں کاغذ کے ساتھ ساتھ سی ڈی پر بھی دستیاب ہونے لگیں تو طلبہ کو ان سے اپنے کمپیوٹر پر خاص طور پر فنکشنل زبان سیکھنے میں بے حد مدد ملے گی۔

لفظ کار کا "متن" ایک سیال (Fluid) حثیت رکھتا ہے۔ اسے جیسے چاہیں کسی بھی ہیئت میں بدل لیں۔ ہر پیراگراف کے ہر جملے کو مختلف زبانوں میں تبدیل کر لیں۔ ہر جملے کو سوالیہ اور منفی میں بدل ڈالیں اور مختلف کھیل/تدبیریں اختیار کر لیں۔ دراصل لفظ کار "ذہن اور صفحے" کو یکجا کر دیتا ہے۔ یہ متحرک اور متعامل رہتا ہے۔ بشرطیکہ کوئی طالب علم واقعی اس سے کچھ سیکھنا چاہے۔

اس مرحلے پر استاد کیا کر سکتا ہے؟ وہ انھیں کمپیوٹر مشین سے متعارف کرانے اور لفظ کاری کے اصول بتانے کے بعد یہ بتا سکتا ہے کہ وہ کاغذ کی بچت کرنے کے ساتھ ساتھ قلم گھسانے اور خوش خط لکھنے کی زحمت سے بھی بچ جائیں گے اور سکرین ہی پر مختلف مشقیں انجام دے سکیں گے۔

خاص طور پر مضمون نگاری اور انشاء پردازی میں اپنے الل ٹپ اور بکھرے ہوئے خیالات کو ٹائپ کر کے / یا کرا کے بعد ازاں کسی منطقی ترتیب میں لا کر ایک مبسوط تحریر پیش کر سکتے ہیں۔ یا کام انفرادی کے علاوہ گروہی طریقے سے بھی انجام دیا جا سکتا ہے۔ اگر سکول میں صرف ایک کمپیوٹر مہیا ہو تو مختلف گروہوں کو مختلف اوقات تقسیم یا الاٹ کیے جا سکتے ہیں۔ لیکن اگر جماعت میں استاد اور طلبہ کو کمپیوٹر پر بنی ہوئی مشقیں، سوالات، پرچے، اور سرگرمیاں طلبہ کو مہیا کر سکتا ہے جو ان کے جوابات تحریر کر سکتے ہیں۔ خاص طور پر اعلیٰ ثانوی سطح پر اب کمپیوٹر کو اُردو کی جماعت میں داخل ہو جانا چاہیے اور اس کے لیے اسباق لکھے جانے چاہئیں۔

اگرچہ اُردو میں ابھی ''تدوینِ متن'' (text editing) کی سہولت میسر نہیں جو الفاظ کے ہجے میں، تراکیب اور محاورں کی درستی سکرین پر پیش کر سکے لیکن استاد ایسی مشقیں کمپوز کر کے طلبہ میں ان کی فلاپی ڈسک تقسیم کر سکتا ہے۔

اگرچہ اُردو میں ابھی اشاریہ سازی، الفائی ترتیب الفاظ اور دیگر ایسی کئی سہولیتیں میسر نہیں تاہم طلبہ کو ''دیکھو''، ''چوا ور ترتیب دو'' کے انداز سے الفاظ کو الٹا الٹا کر الفبائی ترتیب اور ایسی کئی سرگرمیوں میں فعال کیا جا سکتا ہے۔ مختصر یہ کہ لفظ کار کے ذریعے معروف سے مجہول جملے، سادہ سے مرکب جملے، سوالیہ سے منفی جملے، صرفی ونحوی تقسیم سکھائی جا سکتی ہے۔ تہنیتی کارڈ، خط، درخواستیں، ان کی مختلف صورتیں، اور پیغامات کے کینڈے (Genres) سکھائے جا سکتے ہیں۔ متن کا تجزیہ کیا جا سکتا ہے۔ قواعدی تجزیہ کیا جا سکتا ہے۔ اس میں سے اسماء افعال، مصادر، حروف الگ الگ کیے جا سکتے ہیں۔ اور سب سے بڑی بات کلوز ٹیسٹ تیار کیا اور دیا جا سکتا ہے۔ تلخیص کرنا سکھایا جا سکتا ہے۔ اُردو کی عملی / فنکشنل صورتیں سکھائی جا سکتی ہیں۔ کہانیاں اور معمے غرض یہ کہ امکانات کی ایک وسیع دنیا موجود ہے جو کمپیوٹر کے ذریعے اُردو کی تدریس کا مستقبل روشن کر رہی ہے۔

''اخبارِ اُردو'' میں ایسے کئی امکانات کا ذکر کیا گیا ہے۔ جو اُردو اطلاعیات کے افق پر روشن ہو رہے ہیں۔ ان میں اُردو کی پہلی ڈیجیٹل یا برقیاتی کتاب بھی شامل ہے۔ ضرورت اس امر کی ہے کہ اُردو کے اساتذہ خوابِ گراں سے بیدار ہوں اور پوری جدتِ طبع اور ذہانت کے ساتھ اس نئے چیلنج کا مقابلہ کریں جو کمپیوٹر / انٹرنیٹ کی صورت میں دنیا کی زبانوں کو درپیش ہے۔ صورت حال یہ ہے کہ اس وقت دینا کی چار ہزار زبانوں میں سے صرف ۶۶ دفتری / سرکاری استعمال میں ہیں اور ان میں سے بھی صرف ۲۶ کو انٹرنیٹ کے انتخاب پر دیکھا جا سکتا ہے۔ اُردو ان ۲۶ زبانوں میں شامل نہیں ہے۔ جبکہ ہندی، بنگلہ، بلکہ گجراتی اور تامل جیسی زبانیں موجود ہیں لیکن کوئی پاکستانی زبان وہاں نہیں ہے۔ یونی کوڈ Unicode جیسے عالمی ضابطوں میں بھی بھارت کی یہ زبانیں شامل ہیں لیکن اُردو کا رسم الخط نستعلیق نہیں۔ اگر اُردو کے اساتذہ نے اس طرف ابھی سے

توجہ نہ دی تو پھر اُردو واقعی کہیں نہیں ہو گی۔ دیگر پاکستانی زبانوں کا تو پھر سوال ہی پیدا نہیں ہوتا۔

ضروری ہے کہ اُردو کے اساتذہ بھی کمپیوٹر کے میدان عمل میں اتر آئیں اور کمپیوٹر کے ذریعے اُردو کی تدریس کے طریقے وضع کرنا شروع کریں۔ اُردو کے اساتذہ بازار میں سی ڈی کے تیار ہونے اور فروخت کے لیے آنے کا انتظار نہ کریں۔ سکول انتظامیہ تھوڑے سے خرچ سے سافٹ ویئر کمپنیوں سے ایسی سی ڈی لکھوا سکتی ہے جو ان کی فوری تدریسی ضرورتیں پوری کر سکتی ہے۔ اساتذہ کو ایسی اختراعی صلاحتیوں کو بروئے کار لانا چاہیے جن سے ان کی تدریس جدید اور بہتر بنیادوں پر استوار ہو سکے۔

("اخبارِ اُردو"، اسلام آباد، اکتوبر ۲۰۰۲ء)

۹

جدید تدوینِ متن اور اُردو اطلاعیات

متنی تنقید (Textual criticism) عیسائیت کے پیروکاروں اور اس کے نقادوں اور محققوں کی وضع کردہ ایک اصطلاح ہے۔ لیکن یہ جانے بغیر کہ دراصل یہ میدان **بائیبل** کے ''ماخذوں'' کی تفہیم اور تجہیل پر آمادۂ پیکار لوگوں نے عیسائیت کے فروغ اور استرداد کے لیے تخلیق کیا، اُردو میں ادبی کام کرنے والوں نے اسے اُردو تحقیق کے متبادل کے طور پر پیش کرنا شروع کیا اور متنی تدوین کی ضرورتیں پوری کرنے لگے۔ وہاں یہ میدان اب اور آگے بڑھ چکا ہے اور تدوین متن کے بعض جدید ترین تقاضے اعلیٰ متنی تنقید (Higher Textual Criticism) کے نام سے سامنے آنے لگے ہیں۔

اس میدان تحقیق کے لیے ایک اور اصطلاح تنقید تسوید (Redaction criticism) بھی استعمال ہوتی ہے۔ مقصود صحائف پر تحقیق اور ان کے ''مصنفین'' کا سراغ لگانا ہے تاکہ ان کے استناد اور جعل کو ثابت کیا جا سکے۔ ان تمام کاوشوں کا نتیجہ زیادہ تر **بائیبل** کو رد کرنے ہی سے نکلتا ہے۔

بائیبل کی تدوین

بائیبل پر کی گئی تحقیقات کے دو پہلو ہیں۔ پہلا متنی تنقید اور دوسرا اعلیٰ متنی تنقید ہے۔ پہلی کاوش متن کے اصل الفاظ کی دریافت ہے جبکہ ہمارے پاس اس کے اصل الفاظ موجود نہیں۔ دوسری متن کی اصلیت سے متعلق ہے کہ وہ متن حقیقت میں کب وجود میں آیا یا لکھا گیا اور کس نے لکھا؟ اُردو تحقیق کے اس پہلو کہ متن کی تدوین کے بعد متن پر تحقیق انجام دی جائے، پہلے کی نسبت زیادہ کارآمد اور مفید ہے۔ چنانچہ اعلیٰ متنی تنقید کو اُردو تحقیق کا ایک باب بنایا جا سکتا ہے جبکہ متنی تنقید کو صرف لفظی جائزے یا تدوینِ متن کے طور پر لیا جا سکتا ہے۔

اعلیٰ متنی تنقید کے ماہرین کا خیال ہے کہ **بائیبل** محض زبانی روایات کی ایک تحریری صورت ہے اور کسی روح القدس نے انھیں لوگوں پر القا نہیں کیا اور **عہد نامۂ قدیم** بابل میں یہودیوں کی قید (۵۸۶ قبل مسیح) سے پہلے ورطۂ تحریر میں نہیں آیا۔ اگر حضرت موسیٰ کے دیے گئے قوانین اور پہلے پانچ صحائف حضرت موسیٰ کے زمانے میں تحریر نہیں ہوئے تھے تو یہ کب تحریر ہوئے تھے۔ اسی طرح **اناجیل** اگر حضرت عیسیٰ کے زمانے میں

نہیں لکھی گئی تھیں اور عرصہ بعد محض زبانی روایات کے مرتبین نے انھیں وجود بخشا تو کیا یہ انھی مصنفین نے تحریر کی تھیں، جن کے نام ''یوحنا''، ''متی'' وغیرہ دیے گئے ہیں یا کسی اور نے ان کی روایات کو قلم بند کیا تھا؟ یہ تمام محققین انھیں الہامی کتب ماننے سے بنیادی طور پر انکاری ہوتے ہیں۔

لغات اور انسائیکلوپیڈیا میں اعلیٰ متنی تنقید کی تعریف کچھ اس طرح سے کی گئی ہے۔

''**بائبل** پر تنقید کی ایک قسم جو متنی تنقید سے مختلف ہے۔ اس کا مقصود اعترافی یا عقائداتی الٰہیات سے قطع نظر **بائبل** کے متن کی تشریح کرنا ہے۔ یہ **بائبل** کے متن پر وہی اصول لاگو کرتی ہے جو سائنسی اور تاریخی طریق سے اخذ ہوتے ہیں۔ یہ زیادہ تر داخلی شہادت پر مبنی ہے اور لسانیاتی کوائف اور آثاریات سے حاصل شدہ شواہد بھی استعمال میں آتے ہیں۔ بنیادی سوالات یہ ہیں کہ استناد کی دریافت، قرائن کی ترتیب و تنظیم کو متن کے مختلف ماخذوں سے جانچا جائے اور مصنفین کے تشخص اور نیتوں کو بھی ملحوظ رکھا جائے۔ اس کا آغاز جرمن سکالر جوہاں سلوموسیملر (۹۱۔۱۷۲۵ء) سے ہوا تھا''

اعلیٰ متنی تنقید کا ایک بنیادی اصول ایک سکالر چارلس اے برگس (Briggs) (۱۸۴۱ء) سے سامنے آتا ہے۔ لیکن ۱۸۸۶ء میں یہ ایک سکالر گراف ولہوزن (Graf Wellhausen) کا فرضیہ کہلایا۔ اس کے نزدیک **تورات** چار ماخذوں سے وجود میں آئی۔ جنھیں ایک محرِّر نو (Redactor) **R** نے مرتب کیا۔

J : یہووائی (انسان نما خدا) ایک شخص تھا جو عبرانی کی ابتدائی صورت استعمال کرتا تھا، جو **P** استعمال نہیں کرتا تھا۔

F: ایلوہی (انسان نما خدا) ایک شخص تھا جو عبرانی کے بعض قواعد کے مطابق ایلوہیم کہلاتا تھا، بعد میں یہوِیہ (Yahweh) تھا جس کی بنیادی دلچسپی بائبل اسرائیل اور شیلوہ کی پیشوائی تھا۔ اس کا اسلوب بھی **P** سے قدیم تر لیکن ذرا بہتر تھا۔

P: پیشواؤں کا ماخذ ایک دور اور بے رحم خدا جسے کبھی ایلوہیم اور کبھی ایل شدائی کہا گیا۔ یہ جزوی طور پر **J** اور **E** کی نقل تھا، لیکن تفاصیل کو اپنی ضرورت کے مطابق ادلتا بدلتا تھا۔ اس کی دلچسپی کا محور ہارونی پیشوائی اور شاہ حزقیاہ تھے۔ یہ نچلے درجے کا ادبی اسلوب استعمال کرتا تھا اور اسے فہرستوں اور تاریخوں سے دلچسپی تھی۔

D: یہ ذونامی (Deuteronomy) میں دلچسپی رکھتا تھا۔ اس نے یشوع، قضاۃ ۱، ۲، سموئیل ۱، ۲ اور سلاطین ۱، ۲ تحریر کیں۔ اسے شیلوہ کی پیشوائی اور بادشاہ یسعیاہ سے دلچسپی تھی۔ وہ **P** کی طرح کی عبرانی زبان استعمال کرتا تھا لیکن دوسرے ادبی اسلوب میں۔

اب انھیں خدا مان لیں یا اشخاص، عہد مان لیں یا افراد JEPD بہرحال ایک گروہ ہے جس نے مل

کر **بائیبل** تحریر کی۔

اعلیٰ متنی تنقید کا استعمال **قرآن مجید** پر بھی ہونے لگا ہے اور اس کے مختلف نسخوں اور ورژنوں کو دیکھا جانے لگا ہے اور یہ اصول ادبی تحقیق میں بھی استعمال ہوتے ہیں، جن میں نظریہ معلومات یا اطلاعات بھی زیر استعمال ہے۔

اعلیٰ متنی تنقید ایک تحقیقی ڈیزائن وضع کرتی ہے جو کسی نہ کسی ضرورت پر مبنی ہوتی ہے۔ پھر ایک موزوں طریق تحقیق کو استعمال کرتی ہے تاکہ تحقیقی ڈیزائن کے لیے مطلوبہ معلومات فراہم ہوسکیں۔ نتائج کا تحقیقی ڈیزائن کے مطابق تجزیہ کیا جاتا ہے اور پھر اس کے نتائج شائع کیے جاتے ہیں۔ یہ تمام مراحل کسی معیار کے مطابق وضع کیے جاتے ہیں۔ معیارات کی تخلیق کے رہنما اصول تاریخی دستاویز کا رہی سے اخذ کیے جاسکتے ہیں جس میں مقاصد سے لے کر سفارشات درج کرنے تک کے نکات بیان ہوتے ہیں۔

اعلیٰ متنی تنقید کسی کتاب (**بائیبل**) کے مختلف نسخوں، ترجموں اور ادوار میں پیش کیے گئے متنوں میں غیر ارادی تبدیلیوں کو بھی ملحوظ رکھتی ہے اور اسے ایک بہت بڑا تحقیقی اصول مانتی ہے۔ اس کے علاوہ ہجوں اور قواعد کے لحاظ سے تبدیلیوں پر بھی نظر رکھتی ہے۔ بیری آر لنسکی نے اس کے نکات واضح کیے ہیں۔

سوال یہ پیدا ہوتا ہے کہ اعلیٰ متنی تنقید کا واقعی کوئی تعلق سکالرشپ یا علمیت سے ہے یا یہ محض ایک تکنیک ہے۔ ڈاکٹر رابرٹ ڈک ولسن اس کے بارے میں کہتا ہے کہ اگر **بائیبل** شکوک وشبہات کا شکار ہوئی ہے تو ایمان رکھنے والوں کو بھی اسی تکنیک اور اصول تحقیق کو اپنانا ہوگا۔ انھیں بھی تاریخ، قرائن، شواہد ہی کی روشنی میں اس کا جواب تلاش کرنا ہوگا جیسے ان سوالوں کا جواب کہ فارسی الفاظ **بائیبل** سے غائب کیوں ہیں؟ یونانی الفاظ کیوں شامل ہیں؟ دانیال نے فارسی الفاظ کیوں استعمال کیے؟ عبرانی میں آرامی کیوں؟ عبرانی کے ادبی اسالیب کیا تھے؟ وغیرہ۔

اغلاط کا مسئلہ

متنی تنقید کے دو اصول (۱) خارجی شواہد (۲) داخلی شواہد اور مخطوطوں کا امتزاج ہیں۔ اعلیٰ متنی تنقید میں دستاویزی فرضیہ کام کرتا ہے جس کا ذکر ہم تورات کے حوالے سے کر آئے ہیں۔ دانستہ یا غیر دانستہ اغلاط کا ایک ذکر ڈاکٹر خلیق انجم کے حوالے سے ہم کر آئے ہیں۔ اب کلین (Klein) کے حوالے سے دیکھتے ہیں:

(الف) غیر دانستہ یا غیر ارادی

۱۔ تجنیس حرفی اور تجنیس شکل حرفی کا التباس (Similar letters)

۲۔ تجنیس صوتی اور تجنیس صوتی حروفی کا التباس (Sound alike)

۳۔ ہم ختمہ حذف یعنی یکساں ختمہ پر حذفِ حروف(Homoeoteleuton)

۴۔ ہم آغاز حذف یعنی یکساں آغاز پر حذفِ حروف(Homoeoarchton)

۵۔ ہم نگاری یا دو حروف یا الفاظ کو ایک سا تحریر کرنا(Haplography)

۶۔ مکرر نگاری یا لفظ یا حرف مکرر درج کرنا(Dittography)

۷۔ لفظوں میں غلط فاصل ہونا یعنی الفاظ کا مل جانا

۸۔ غلط حروفِ علت درج کرنا(Incorrect vocalization)

۹۔ ماورائے دعویٰ یعنی الفاظ یا حروف الٹ پلٹ لکھنا(Meta thesis)

۱۰۔ غلط مترادفات دینا(Substitution of synonyms)

۱۱۔ کسی پیرے کے الفاظ ذرا سے بدل کر آگے لکھنا

۱۲۔ متن میں حواشی یا تشریح شامل کر لینا

(ب) دانستہ یا ارادی

۱۔ ہجوں یا قواعد میں تبدیلی

۲۔ ہم آہنگی یا ملتے جلتے الفاظ اضافہ کرنا(Harmonization)

۳۔ دو مختلف نسخے یا قرأتیں شامل کرنا(Conflation of variants)

۴۔ نام اور کنیت شامل کر لینا

۵۔ فاعل اور مفعول کی تشریح کرنا

۶۔ کسی اور پیرا سے معلومات کا اضافہ کرنا

۷۔ مشکل مقامات حذف کر دینا

۸۔ غریب اور نادر الفاظ نکال کر بدل دینا

ان امور کی تشریحات بھی موجود ہیں جن کے لیے اُردو میں مثالیں اور متون تلاش کرنا درکار ہے۔ انگریزی میں کلین ہی سے بائیبل کی چند مثالیں ملاحظہ ہوں۔

Unintentional Changes

1. Confusion of similar letters.I Samuel 14:47:
MT: "He pronounced (them) wicked" (y r sh y ')
LXX: "He was victorious" (y w sh ' -- reading a w for an r)
2. Confusion of words that sound alike. I Samuel 28:2

MT: "you" ('attah)

LXX: "now" (apparently reading 'attah)

3. Omission because of similar endings(homoeoteleuton.

I Samuel 13:15:

MT: "And Samuel arose and set out from Gilgal to Gibeah of Benjamin"

LXX: "And Samuel arose and set out from Gilgal-- and went on his way; but the rest of the people went up after Saul to meet the soldiers. Then they came from Gilgal--to Gibeah of Benjamin."

4. Omission because of a similar beginning (homoeoarchton).

5. Haplography or single writing. I Samuel 17:46 has such a case:

LXX: "I will leave your corpses and the corpses of the Philistine army" (the words apparently coming from consonants p g r k)

MT For the words in italics the MT only has one p g r.

6. Dittography or double-writing. II Samuel 6:3-4:

"And they made the ark of God ride on a new cart, and they took it away from the house of Abinadab which is on the hill. Uzzah and Ahio, sons of Abinadab, guided the--new cart, and they took it away from the house of Abinadab which is on the hill."

7. Incorrect word division.

MT: "And he built the city"(I Chron. 11:8 [w y b n h ' y r])

LXX: "And he built it a city" (II Sam. 5:9 [w y b n h ' yr])

8. Incorrect vocalization. Psalm 130:4

MT: "there is forgiveness that you might be feared" (tiwware')

LXX: "law" (the translator saw the consonants and assumed it was the common noun t w r ' [tora], rather than a very rare, irregular verb he probably did not know).

9. Transposition of words or letters (metathesis).Samuel 17:39):

MT: "and he endeavored unsuccessfully" (w y ' l)--an awkward reading!

LXX: "and he exerted himself" (apparently reading w y l ')

10. Substitution of synonyms. I Samuel 10:25 has:

MT: "each man to his home"

LXX: "each man to his place" (also in 4QSama)

11. I Samuel 12:15

MT: "the hand of Yahweh will be against you and your fathers."

The reading "fathers" is difficult. LXXL has "your king." S. R.

Driver suggests that the frequent use of "fathers" in verses 6-8 may have led to the change accidently.

12. Mistaken inclusion of marginal comments into the text. Isaiah 24:4:

MT: "the heights with the land (mourn)"

1QIsa "the heights of the land (mourn)"

Intentional Changes

1. Changes in spelling or grammar. Modern translations often do this as well, occasionally putting the additions in italics, but not always.

2. Harmonizations.I Samuel 20:5 verses 34-35 tells that David hid for three days.

MT: "Let me hide in the open country until the third evening."

LXX: "Let me hide in the country until evening."

3. Conflation of variant readings.Ezekiel 1:20, the italicized words are missing in some Hebrew manuscripts, LXX, and Syriac.

MT: "Wherever the spirit wanted to go, they went, wherever the spirit wanted to go, and the wheels rose along with them."

4. Filling out names and epithets.II Samuel 3:3a

MT: "Chileab of Abigail (k l ' b l ' b y g l) the widow of Nabal the Carmelite"

LXX: "Dalouia the son of Abigaia the Carmelitess"

5. Supplying subjects and objects. "if LXX and MT differ in respect of a subject, it is probable that the original text had neither" (see Driver, Notes on the Hebrew Text, p. lxii).

6. Expansion from parallel passages.

7. Removal of difficult expressions.Job 1:5, 11 and 2:5, 9, the expression "curse God" was the original. The expression was offensive, and so was changed for the euphemism "bless God"--although the scribes knew it was "curse God."

8. Replacement of rare words with more common ones. Isaiah 39:1

MT: "[Hezekiah] became well" (w y kh z q)

1QIsa "[Hezekiah] became well" (w y kh y h)

علمِ زبان(Philology)انسانیاتی شعبے کی ایک شاخ ہے، جس کا اطلاق ادبی متون پر کیا جاتا ہے جو روایتی طور پر یہ ادنیٰ واعلیٰ تنقید پر منقسم ہے۔ اس کا بنیادی مقصود مصنف کے اصل متن کی بازتخلیق ہے، اعلیٰ صورت میں مصنف کے کام، اسلوب اور متن کی حقیّت کو ثابت کرنا ہے۔ نظریۂ اطلاعات و معلومات (Information Theory) سے حاصل کردہ طریق تحقیق کا اطلاق اسے دونوں میدانوں کے بعض

پہلوؤں سے ہم آہنگ کرنے کا باعث بنتا ہے۔متنی تنقید اور استحسان دونوں میدانوں میں اس کا تجربہ انگریزی، اطالوی اور دوسری عالمی زبانوں میں بے حد حوصلہ افزا رہا ہے۔متنی تنقید میں تین مختلف روایات کی جانچ روایتی طریقے کے نتائج سے ہم آہنگ رہی ہے جس میں بہت زیادہ وقت درکار ہوتا ہے۔ یہ تجربہ تیرھویں صدی کے اطالوی شعرا اور مکاتبِ فکر پر کیا گیا جس سے ظاہر ہوتا ہے کہ متون کو ایک دوسرے کے قریب لا کر جانچنا ممکن ہے۔مزید برآں اس سے گمنام تحریروں کا دور متعین کرنے اور ان کا استحسان کرنے میں سہولت حاصل ہو جاتی ہے۔

نظریۂ اطلاعات

نظریۂ اطلاعات یا معلومات کیا ہے۔اس کی ایک تعریف یوں کی گئی ہے۔

> ''نظریۂ اطلاعات اطلاقی ریاضی اور انجینئرنگ کی ایک شاخ ہے جس میں معلومات یا اطلاعات کی مقدار پیمائی کی جاتی ہے......دیگر امور کے علاوہ یہ ابلاغ کے مطالعے کے لیے بھی استعمال ہوتی ہے''۔

عام طور پر خارجی اور داخلی اصولوں اور شواہد میں خاصا امتیاز کیا جاتا ہے۔ خارجی اصولوں سے دوسرے مصنفین کی شہادتیں جمع کی جاتی ہیں جن میں تاریخی، ثقافتی اور سوانحی حوالے متون میں تلاش کیے جاتے ہیں اور ماخذوں کا تجزیہ کیا جاتا ہے۔اس کے برعکس داخلی اصولوں میں علم زبان کے شواہد تلاش کیے جاتے ہیں، جن میں عروض، اسلوب، لفظیات وغیرہ سے بحث ہوتی ہے۔ یہ بھی بجا کہ اسلوبیات میں اسلوب پیمائی کے عددی اور مقداری طریقے بھی استعمال ہوتے ہیں اور طویل متن کے مختلف ورژن باہمی تقابل سے جانچے جاتے ہیں۔اس کی خاصی تفاصیل بیلی نے فراہم کر رکھی ہیں، جن میں ساخت، تعدد اور مصنف کے شعوری و لاشعوری الفاظ و انداز کو سمجھا جاتا ہے۔ خارجی و داخلی شہادتیں زبان میں الفاظ کے امکانات کا ایک جزو ہیں۔

نظریۂ اطلاعات کے دائرہ کار کی بنیاد ''زبان'' ہی ہے۔ایک عمدہ زبان میں دو بڑی خصوصیات پائی جاتی ہے۔اوّل، عام مشترک الفاظ، انھیں خاص اور غیر مشترک الفاظ کی نسبت چھوٹا ہونا چاہیے جیسے ''میں''، ''تو''، ''وہ'' تاکہ جملے طویل نہ ہوں۔دوم، اگر جملے کا کوئی حصہ کسی شور/ مداخلت کی وجہ سے سنا/ پڑھا نہ جا سکے تو سامع یا قاری مطلب یا مفہوم ضرور سمجھ لے۔

نظریۂ اطلاعات ۱۹۴۸ء میں کلاڈ شینون (Shannon) نے اپنے کام ''ابلاغ کا ریاضیاتی نظریہ'' میں پیش کیا تھا۔اس میں امکانات اور شماریات کو یک جا کیا گیا تھا۔الگورزم کو اس کا حسابی آلہ بنایا گیا جو دو متغیرات کے درمیان نسبت کی تلاش کا طریقہ ہے۔

اطلاعات کی پیمایش کا ایک ذریعہ حروف ہیں اور اطلاعیات کے میدان میں انھیں ہم بٹ(Bit) میں ناپ سکتے ہیں جو کمپیوٹر میں معلومات ریکارڈ کرنے کی اکائی ہے یعنی صفر یا ایک۔ اگر تین بٹ ہوں تو ان کے جوڑ آٹھ طریقوں سے سامنے آتے ہیں، (جیسے 000، 001، 010، 011، 100، 101، 110، 111)۔ انھیں ہم الگورزم کی زبان میں Log8 کہیں گے۔ اگر آپ کسی سکے کو دس لاکھ بار اچھالیں اور نتائج کا سلسلہ درج کرتے جائیں اور اس کا ابلاغ کرنا چاہیں تو آپ کو کتنے بٹ درکار ہوں گے؟ اس مسئلے کو یوں حل کیا جائے گا۔ سکے کے دو رخ ہوتے ہیں جو سامنے آنے کے ایک جیسے امکانات رکھتے ہیں یعنی پچاس پچاس فی صد۔ ہر ایک امکان کے ابلاغ کے لیے ایک بٹ درکار ہے۔ دس لاکھ بار اچھالنے کے ابلاغ کے لیے دس لاکھ بٹ درکار ہوں گے۔ فرض کریں کہ سکے کا ایک رخ لازماً صرف ۲۵ فی صد بار سامنے آتا ہے تو سارا ابلاغ ۳۰۰،۸۱۱ بٹ میں منتقل ہو سکے گا۔ اس کا مطلب ہے کہ سکے کے ہر اچھال کے لیے ۸۱۱۳ء بٹ درکار ہوں گے۔ یہ کیا ہوا؟ ایک سے بھی کم بٹ؟ اگر یہ لازماً ۲۵ فی صد اور کم ہو جائے تو بٹ کی یہ ضرورت بھی اور کم ہوتی جائے گی۔ امکانات کا یہ معاملہ ارسطو کے نظریہ کے تحت ہوتا جاتا ہے۔ یعنی امکانات کی تعداد بڑھتی اور قوت کم ہوتی جاتی ہے جیسے:

۱/۲ ، ۱/۴ ، ۱/۸ ، ۱/۱۶ ، ۱/۳۲ ، ۱/۶۴............

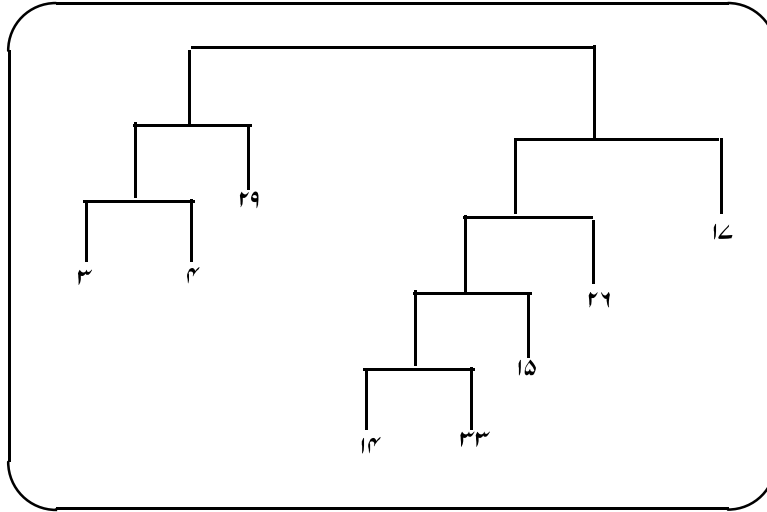
اس سارے کے مجموعے کو یونانی حرف سگما Σ کی علامت سے ظاہر کیا جاتا ہے اور الگورزم کے استعمال سے، جس میں P امکانات کے لیے استعمال ہوتا ہے فارمولا کچھ یوں ہوگا:

$$- \Sigma \ p(x) \log p(x)$$

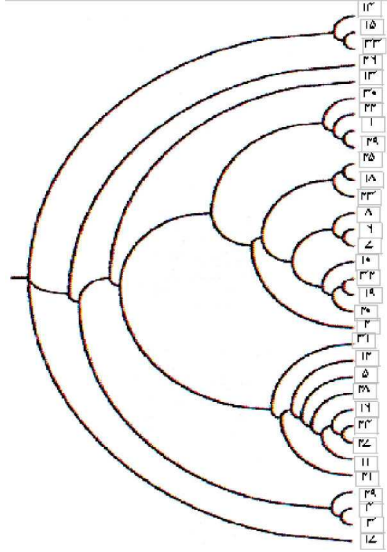
ادبی امور اور متنی تنقید کے لیے ہمیں زیادہ ریاضیاتی اور شماریاتی فارمولوں سے تو بحث نہیں لیکن ان کی بنیاد سمجھ لینا بہت ضروری ہے۔ امکانات کی دنیا کو ہم کسی لفظ کے بہت سے امکانات یا کسی کتاب کے بہت سے نسخوں یا اشاعتوں میں وقوع پذیری کو ملحوظ رکھ کر جان سکتے ہیں، جیسا کہ متنی تنقید کے اصولوں میں ماخذوں کے شجرے کی بات ہوئی تھی، جس میں اغلاط اور اشتراکات کے امکانات سامنے آتے ہیں۔

نسخوں کا کمپیوٹری شجرہ

طباعت سے پہلے کتابیں ہاتھ سے لکھی جاتی تھیں اور مختلف نسخوں میں اختلافات راہ پا جاتے تھے۔ متنی تدوین انھی اختلافات کا جائزہ لینے اور درست نسخہ مرتب کرنے کا علم ہے۔ مسئلہ یہ ہے کہ ایسا صرف قلمی نسخوں ہی میں نہیں ہوتا مطبوعہ نسخوں میں بھی ایسا ہوتا ہے۔ جیسا کہ فسانۂ عجائب کی ۱۸ کے قریب اشاعتیں مصنف کی زندگی ہی میں ہوئی تھیں۔ بعد کی کئی اشاعتوں کو ملا کر صحیح نسخہ مرتب کرنے کے لیے ان اٹھارہ کے اٹھارہ نسخوں کے ساتھ دیگر کا تقابل کرنا ضروری ہوگا۔ متنی تنقید کے قدیم طریقے میں مرتب یا تحقیق کار ایک نسخے

کو بنیاد بناتا ہے اور مختلف نسخوں کے ساتھ تقابل کرتا ہے لیکن اعلیٰ متنی تنقید کے اطلاعاتی طریقے میں متغیرات کو غلطی کے انداز تلاش کرنے کے لیے زیر مطالعہ لاتا ہے جس سے متن کی تاریخ مرتب ہوتی ہے۔ اس کے لیے اس کو نسخوں یا ضابطوں کا شجرہ (Stemma Codicum) وضع کرنا پڑتا ہے۔ نظریۂ اطلاعات میں کلیدی عنصر دو مسودات کے مابین ''فاصلے'' کی تعریف کرنا اور کوائف کے محولہ بالا امکانات کے درمیان ربط پیدا کرنا ہے۔ اس طرح وضع کردہ شجرے کو دوسرے دو مسودات کے درمیان پیدا شدہ شجرے کے ساتھ موازن کیا جاتا ہے۔ شجرے کی یہ تکنیک ہم اسماء الرجال، لسانی شاخوں، انواع واقسام کے رابطوں، تعلقات اور نسبتوں کی تشریح کے لیے استعمال کرتے ہیں۔ شجرہ بنیادی طور پر حیاتیات میں استعمال ہوتا ہے اور وہاں اس کام کے لیے سافٹ ویر تیار ہو چکے ہیں۔ ادبی متون کے تقابلی مطالعے کے لیے اس سافٹ ویر کا پہلا استعمال Canterbury Tales Project میں ۸۴ قلمی نسخوں اور ۴ مطبوعہ نسخوں کے تقابل کے لیے استعمال کیا گیا۔ ان کے متون کمپیوٹر میں داخل کیے گئے اور تمام اختلافات سافٹ ویر کی مدد سے حاصل کیے گئے۔ بعدازاں ان نسخوں کو مشترک خصوصیات کی بنا پر گروہ بند کیا گیا۔ یہ سافٹ ویر ''تلاشِ لفظ'' اور ''اشاریہ سازی'' کے سافٹ ویر/ایڈیٹر سے ذرا مختلف نوعیت کا ہوتا ہے۔ اس کے استعمال کے لیے اس کی پیچیدگی کو سمجھنا ضروری نہیں۔ یہ سافٹ ویر لسانی اختلافوں اور مشابہتوں، زبان کی تاریخ اور لسانی ثقافت کو سمجھنے کے لیے مفید ثابت ہوئے ہیں۔ ذیل کا نقشہ مختلف نسخوں کے اختلاف کو سمجھنے میں مدد دے سکتا ہے۔

۳۳ نسخوں میں سے دو نسخوں نمبر ۳، ۴ میں ایک جیسے الفاظ ہیں لیکن وہ دوسروں سے مختلف ہیں۔ نمبر ۳، ۴ کے بعض اشتراکات نمبر ۲۹ کے ساتھ مشترک ہیں لیکن باقی نسخوں سے مختلف ہیں۔ اسی طرح نمبر ۱۴، ۳۳ بعض مقامات پر مشترک ہیں، پھر وہ دونوں مل کر نمبر ۱۵ کے ساتھ مشترک ہیں۔ وہ تینوں مل کر نمبر ۲۶ کے ساتھ

مشترک ہیں اور پھر یہ سب نمبر ۱۷ کے ساتھ مشترک ہیں۔

بینے وغیرہ نے ۳۴ نسخوں کے تقابل سے جو شجرہ بنایا وہ ظاہر کرتا ہے کہ نمبر ۱۷ اقدیم ترین نسخہ ہے ۔ نسخوں یا ضابطوں کا شجرہ ملاحظہ ہو۔

یہ تکنیکیں ہمیں بتاتی ہیں کہ نسخوں کا شجرہ اور خاندان بنا کر ہم ان کی قدامت کا اندازہ لگا سکتے ہیں اور اپنی فنی تدوین کو بھی تحقیق کا سائنسی مقام عطا کر سکتے ہیں۔ اگر چہ فی الوقت یہ کام پیچیدہ، ریاضیاتی یا شماریاتی اور کمپیوٹر سافٹ دیر سے متعلق معلوم ہوتا ہے لیکن یہ نسبتاً تیز رفتار آلہ ہے اور متن کی حقیقت معلوم کرنے کے لیے بے حد مفید ہے۔ اُردو میں ایسے سافٹ ویر وضع ہو سکتے ہیں بشرطیکہ جامعاتی شعبہ ہائے اُردو متنی تنقید پر مبنی تحقیق کو سائنسی بنانے کا ارادہ کر لیں۔ اس مقصد کے لیے نظریۂ اطلاعات کو سمجھنا اور اس کا تحقیق میں اطلاق کرنا ضروری ہے۔ یہ روایتی لسانی طریقوں سے زیادہ مستند اور بہتر ہے۔

اعلیٰ متنی تنقید اور نظریۂ اطلاعات کا استعمال الفاظ کی طوالت، ان کے ہجا، سابقوں اور لاحقوں کی پیمایش، تعدد، اسلوبیاتی انداز، مترادفات اور متضادات، متلازمات وغیرہ کے استعمال کے حوالے ہی سے ہوتا ہے۔

نظریۂ اطلاعات، شجروں، لفظوں، عبارتوں کے تقابل کو جاننے کے لیے کچھ اہل تحقیق کو اس میدان میں بھی غوطہ زن ہونے کی ضرورت ہوگی اور چند افراد کو سائنسی ذہن کے ساتھ ادبیات میں کام کرنا ہوگا تب کہیں جا کر اُردو کا شعبہ سائنس کے میدان میں اپنا مقام پیدا کر سکے گا جیسا کہ انگریزی ادبیات نے پیدا کر رکھا ہے۔ سوال یہ ہے کہ کیا ہم خود کو اس کے لیے تیار کرنا چاہتے ہیں؟ کیا واقعی ہم اُردو کو جدید دورِ اطلاعیات

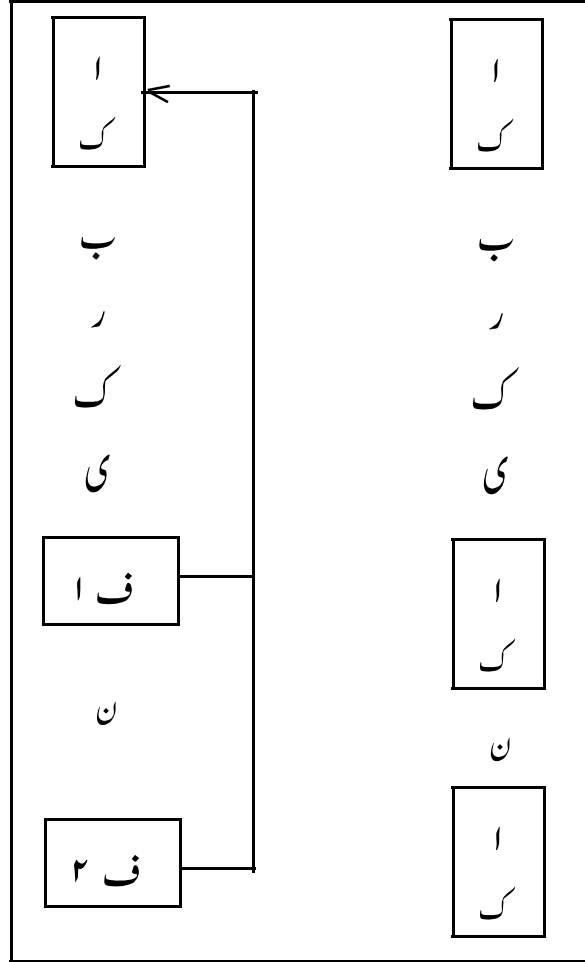
میں قائم رکھنا چاہتے ہیں؟

کتابی نسخوں کی تدوین کو سائنسی تحقیق کا جزو ماننے میں علمائے تحقیق کو ہمیشہ تامل رہا ہے جبکہ اعلیٰ متنی تنقید میں نسخہ پیمائی کو سائنسی طریقوں سے بھی روبہ عمل لایا جاسکتا ہے۔ شرط صرف سائنسی ذہنیت اختیار کرنے کی ہے۔ دورِ جدید میں یہ کام کمپیوٹر کی مدد سے انجام دیا جاسکتا ہے۔ اطلاعیات کا میدان اس کے لیے خاصا سرگرم ہے۔ بنیادی اصول یہ ہے کہ دونسخوں کو باہم جانچنے کے لیے اطلاعیات کا علم اور تکنیک استعمال کی جائے اور نظریۂ اطلاعات کے سافٹ ویر کو استعمال کر کے اس سے حاصل شدہ لفظیات کے ''فاصلے'' کا تجزیہ کیا جائے۔ نظریۂ اطلاعات کا یہ انداز شینون جیسے محقق نے برقیاتی ابلاغ کے حوالے سے پیش کیا تھا۔ آج یہ نظریہ کمپیوٹر سائنس، موبائل فون، خفیہ نگاری، حیاتیات اور طبیعیات کے علاوہ ابلاغ اور ادبیات کے لیے بھی استعمال ہو رہا ہے۔ یہ اکیسویں صدی کی پیداوار اور تحفہ ہے۔

ڈاکٹر خلیق انجم سے لے کر ڈاکٹر تنویر علی، رشید حسن خان اور پھر ڈاکٹر گیان چند تک ہمارے اکثر قدیم اور روایتی محققین ہمیں نسخوں کے تنقیدی ایڈیشن شائع کرنے تک محدود کرتے رہے اور اسی کام کو تحقیق قرار دینے پر بضد تھے۔ انکشافِ متن (Heuristics) اور تنقیدِ متون (Recension) کو تمسیخ (Corruption) کے جس عمل کا سامنا کرنا پڑتا ہے، اس کے لیے اصول بھی وضع ہوئے اور نسخوں کی تقدیم و تاخیر اور گروہ بندی کے کئی طریقے بھی سامنے آئے۔ اختلافِ نسخہ جات معلوم کرنے کی بھی کئی تکنیکیں وضع کرنے کی کوشش کی گئی ہے لیکن ہمارے اکثر تدوین کار کسی نسخے کی تدوین کے بعد ان سب باتوں کو درج کرنا تحقیق کی شان سمجھتے اور یوں ''محقق'' بننے کا عمل پورا کرتے۔ جدید سائنسی تحقیق میں یہ سب کچھ ''گردِ راہ'' ہے جسے ''منزل'' پر موجود نہیں ہونا چاہیے مگر اُردو کے مدونین کا کیا کیا جائے کہ وہ اسی کی پیش کش کو تحقیقِ اعظم قرار دینے پر مُصر ہیں۔ اپنے آلات کار اور حساب کاری ہی کو ''کارنامہ'' سمجھتے ہیں۔ ایسے تمام پہلوؤں کو یہ سافٹ ویر سنبھال سکتا ہے۔ نتیجے کے طور پر نسخوں کی قدامت، استناد، الفاظ اور عبارتوں کی صحت اور درست نسخ کی بنیاد حاصل ہوتی ہے۔

کمپیوٹر میں سلکِ حروف کی طوالت کو کم کرنے کے لیے زِپ (Zip) استعمال کی جاتی ہے جس میں دراصل بعض حروف ''غائب'' ہو جاتے ہیں لیکن امکانی طور پر ''معلوم'' ہوتے ہیں اور حسبِ ضرورت حاصل کیے جاسکتے ہیں۔ نسخوں کے تقابل اور تدوین میں اسی سے مدد لی جاتی ہے جو بعض نامعلوم حروف یا الفاظ کو ''زپ'' کے اصول پر ''معلوم'' کی صف میں ڈال دیتے ہیں۔ گویا نامعلوم پیچیدگی کو جاننے کا ایک بہت اچھا آلہ دستیاب ہو گیا ہے۔ اس میں بتکرار آنے والے حروف کے ''امکانات'' کا حساب کر کے انھیں غائب کر دیا جاتا ہے اور محض امکانات کا فارمولا باقی رہ جاتا ہے۔ اگر یہ فارمولا غیر محدود طوالت کی عبارت پر لگائیں تو تمام نامعلوم حروف یا الفاظ سامنے آجائیں گے۔

اسے Lu77 الگورزم کا نام دیا گیا ہے۔ دلچسپ بات یہ ہے کہ یہ سافٹ ویر ساری عبارت میں مکرر حروف/عبارتیں تلاش کرتا ہے۔ تکرار کو محض عدد کے ایک نشان سے ظاہر کرتا ہے اور ''فاصلے'' کا تعین کرتا ہے اور ایک طوالت پیش کرتا ہے جو حروف کی طوالت کو ''ہم تکراری'' ظاہر کرتی ہے۔ مثلاً کسی متن میں کوئی سلسلہ ہم تکرار (ف،۳) سے ظاہر ہوتا ہے۔ یہاں ''ف'' تکراروں کا درمیانی فاصلہ ہے۔ یاد رہے کہ زپ میں سافٹ ویر الفاظ کو نہیں پہچانتا بلکہ حروف کا ایک ''سلسلہ'' سمجھتا ہے۔ ''۳'' کا عدد اس کی تیسری تکرار یا تعدد کو ظاہر کرتا ہے۔ اوسط ''سلسلہ'' یا لفظ زیادہ سے زیادہ دس حروف/بٹ پر مشتمل ہونا چاہیے تو زِپ کے اصول کے مطابق اسے کم جگہ درکار ہوگی۔

کسی بھی عبارت میں صرف حروف ہوتے ہیں جو بار بار مختلف ترتیبوں کے ساتھ سامنے آتے ہیں۔ اگر ایک ترتیب بار بار آ رہی ہو تو گویا اگلی بار وہ حروف لکھنے کی ضرورت نہیں، بس بتا دیں کہ فلاں حروف کا فلاں گروہ پھر آ گیا ہے۔ زِپ اسی اصول کا نام ہے۔ مثلاً ایک جملہ زِپ میں ملاحظہ ہو ''اکبر کی اک ناک''۔

اس خاکے میں ''اک'' کا لفظ دوسری بار آنے پر ''ف ا'' رہ گیا اور تیسری بار آنے پر ''ف ۲''۔ اب اگر کسی ایسی عبارت میں دوسری بار ''اک'' نہ آئے تو اسے معلوم کیا جا سکتا ہے۔ یہاں ایک اور اصطلاح ''قطع اضافی'' (relative entropy) سامنے آتی ہے، جو دو عبارتوں کے درمیان فاصلے''، ''دوری'' یا ''اجنبیت'' کو ناپنے کا پیمانہ ہے۔ مثلاً اگر ہم کوئی اطالوی عبارت انگریزی حروف سے ٹائپ کرنے لگیں تو ہمیں چند اضافی حروف یا ضابطوں کی ضرورت پڑے گی۔ دونوں کا یہ اضافی فرق ''قطع اضافی'' کہلاتا ہے۔ دو نسخوں کے درمیان بُعد یا اجنبیت اس سے معلوم کی جاتی ہے۔ اسے الفاظ یا سلسلوں کے درمیان ''فاصلے'' سے تعبیر نہ کیا جائے۔ ''قطع اضافی'' ناپنے کے لیے حال ہی میں ایک انداز بنی ڈیٹو نے اختیار کیا ہے۔ مقصود یہ ہے کہ دو متون (الف، ب) کے درمیان ''اجنبیت'' معلوم کرنے کے لیے ''قطع اضافی'' کی خصوصیات معلوم کی جائیں۔

کمپیوٹر میں نسخہ ''ب'' کا ''سلسلہ'' بڑھنا شروع ہوتا ہے اور ''ذیلی سلسلے'' معلوم کیے جاتے ہیں جو نسخہ ''الف'' میں موجود تھے۔ یہ ''سلسلے'' اگر ''ب'' میں مکرر آئیں تو ان پر ''نشان'' لگا دیتے ہیں۔ اس سے ہمیں دونوں نسخوں کے درمیان ''قطع باہم'' (cross entropy) معلوم ہو جاتی ہے اور بالآخر ''قطع اضافی'' حاصل ہو جاتی ہے۔ اب ہم دونوں نسخوں کے درمیان موجود ''اجنبیت'' یا ''دوری'' جان لیتے ہیں۔ ایک ''اجنبیت'' مثلاً ''م'' ہو تو اسے تین ضروریات پوری کرنی چاہئیں: اثباتیت، تشاکل، تثلیثی عدم مساوات۔ اگر قطع اضافی آخری دو خاصیتیں پوری نہ بھی کرتی ہو تب بھی کبھی منفیت ظاہر نہیں کرتی۔ تب ہم متشاکل ''قطع اضافی'' تلاش کر کے سلسلوں کے ایسے ہی جوڑے میں عدم تطبیق معلوم کر سکتے ہیں۔

اعلیٰ متنی تنقید اور اطلاعیات

اس اجنبیت یا عدم تطبیق سے ہم کل مجموعے کے ''شجرے'' وضع کر سکتے ہیں اور یوں ایک جال (Matrix) بُن سکتے ہیں، جس کی بنیاد متون کے جوڑوں کے درمیان ''اجنبیت'' ہے اور اس کے پتے مجموعے کے عناصر ہیں۔ یوں ہم درجہ بندی یا صنف بندی تک پہنچ جاتے ہیں۔

اس طریقے کی درجہ بندی کی اس قوت سے فائدہ اٹھا کر اعلیٰ متنی تنقید میں کام لیا جا سکتا ہے اور کسی خاص عہد کے قلمی نسخوں یا مطبوعات کے مختلف ورژنوں میں تقابل ہو سکتا ہے۔ انگریزی میں اس سافٹ ویر سے کام لیا جا رہا ہے اور نسخوں کی قدامت معلوم کی جا رہی ہے۔ خاص طور پر ایک ہی مصنف کے مختلف نسخوں/ ورژنوں میں ''اجنبیت'' کی تلاش کے ذریعے سے اس عہد کے لسانی کینڈے سے ہم آہنگ ہونے کی ''شرائط'' کے حوالے سے استناد حاصل کیا جا سکتا ہے۔ **کینٹر بری ٹیلز** کے علاوہ **ڈیوائن کامیڈی** کے نسخوں پر بھی اسے آزمایا جا چکا ہے۔

اس سافٹ ویر سے متون میں موجود مرتبین کے اضافے اور الحاقات، اغلاط اور تسامحات الگ کیے جا سکتے ہیں۔ خواہ وہ **مثنوی سحر البیان** کے نسخے ہوں یا **ہیر وارث شاہ** کے، الحاقی اشعار تک کو معلوم کیا جا سکتا ہے۔ اغلاطِ کتابت اور اضافات وملحقات کی نشاندہی ہو سکتی ہے۔ کام صرف یہ کرنا ہے کہ word میں تمام متون ٹائپ کر کے کمپیوٹر میں داخل کر دیے جائیں۔ ٹائپ کرنے یا لفظ کاری (word processing) کے اس عمل میں Inpage جیسے سافٹ ویر کام نہیں دے سکتے۔ اس کے لیے اُردو ونڈوز (M.S. Word) یا مقتدرہ کا پاک نستعلیق فانٹ زیرِ عمل لایا جا سکتا ہے۔ دوسرا سافٹ ویر ان عبارتوں کو دس ہزار حروف کی فائلوں/مسلوں میں تقسیم کر کے ان کا تجزیہ کرنا شروع کر دے گا۔ مصنف کی نثری یا شاعرانہ خاصیتوں کا اندراج کیا جاتا ہے اور تمام متون کے جوڑے شجروں کی صورت میں وضع ہونا شروع ہو جاتے ہیں۔ کئی تجربات کرنے کے بعد بنی ڈیٹو کے اصولوں پر عمل درآمد کرتے ہوئے ''جال'' تیار ہوتے ہیں جو متشاکل متون اور شجروں کی صورت میں مکمل تصویر سامنے لاتے ہیں۔ اس سافٹ ویر میں نثر کی نسبت شاعری پر کام کرنا آسان ہوتا ہے۔ ادبی حوالوں سے متنی تنقید (شکل نمبر ۱) اور تکنیکی شجروں میں پیدا شدہ صورت حال (شکل نمبر ۲) کا تقابل کیا جا سکتا ہے جس میں (Devine Comedy) کے متون کی متنی تنقید کا عمل انجام دیا گیا تھا۔ انگریزی کی مثال اس لیے پیش کی جا رہی ہے کیونکہ اُردو میں ابھی ایسا کوئی عمل سامنے نہیں آیا۔

Figure 1. Summarized and simplified phylogenetic tree of the poetry works of 13th century (the length of the branches does not represent the distances): Dante Alighieri Inferno = 17 files; Dante Alighieri Purgatorio = 17 f.; Dante Alighieri Paradiso = 17 f.; Fiore = 12 f.; Detto d'Amore = 2 f.; Brunetto Latini Favolello = 1 f.; Brunetto Latini Tesoretto = 7 f.; Intelligenza = 10 f.; Cino da Pistoia son. = 1 f.; Dante Alighieri Vita Nova canz. - son. = 4 f.; Guido Cavalcanti canz. - son. = 3 f.; Dino Frescobaldi canz. - son. = 2 f.; Gianni Alfani canz. = 1 f.; Lapo Gianni canz. = 2 f.; Cino da Pistoia canz. 1 = 2 f.; Cino da Pistoia canz. 2 = 2 f.; Dante Alighieri canz. 1 = 2 f.; Dante Alighieri son. 1 = 1 f.; Guido Cavalcanti son. = 1 f.; Dante Alighieri son. 2 = 1 f.; Dante Alighieri canz. 2 = 2 f.

Figure 2. Detail of a branch of the phylogenetic tree illustrated on fig. 2: Poesia cort. tosc. 1 = Panuccio dal Bagno canz. - son. (4 files) and Dante da Maiano canz. - son. (5 f.); Poesia cort. tosc. 2 = Carnino

Ghiberti, Bondie Dietaiuti, Neri de' Visdomini, Maestro Francesco, Inghilfredi da Lucca (2 f.), Bonagiunta Orbicciani canz. (2 f.), Pucciandone Martelli; Poesia cort. tosc. 3 = Chiaro Davanzati canz. - son. (8 f.), Rustico Filippi, Noffo Bonaguidi,
Guittone d'Arezzo son.; Poesia cort. tosc. 4 = Guittone d'Arezzo canz. - son. (24 f.), Bacciarone di messer Baccone, Meo Abbracciavacca, Monte Andrea canz. - son. (11 f.), Jacopo da Leona; Scuola siciliana = Re Enzo, Stefano Protonotaro, Pier delle Vigne, Rinaldo d'Aquino, Federico, Ruggieri d'Amici, Giacomino Pugliese (2 f.), Ruggierone da Palermo, Tommaso di Sasso, Giacomo da Lentini canz. (3 f.), Mazzeo di Ricco, Jacopo Mostacci, Guido delle Colonne; Poesia cort. emiliana = Tommaso da Faenza and Onesto da Bologna; Stil novo = Guido Orlandi, Dino Compagni; Poesia "realistica tosc." 1 = Cecco Angiolieri son. (6 f.), Meo dei Tolomei (2 f.); Poesia "realistica tosc." 2 = Cenne da la Chitarra, Folgore da San Gimignano (2 f.), Muscia da Siena, Cecco Angiolieri son. dubbi (2 f.). Moreover: Chiaro Davanzati caz. - son. = 6 f.; Amico di Dante canz. - son. = 3 f.; Rustico Filippi son. (2 f.); Mareamoroso (2 f.); when the file number is not specified it means that there's only 1 file.

نامعلوم مصنفوں کے کاموں/نسخوں کو بھی اس کے ذریعے سے جانچا جا سکتا ہے اور اس کا عہد تصنیف معلوم ہو سکتا ہے۔ اُردو کا کوائفیہ (Database) جتنا بڑھتا جائے مختلف مصنفوں کے محضر اور طرز بیان، کینڈوں اور اسلوبوں سے تقابل اتنا ہی آسان ہوتا جائے گا۔ لغات نویسی آسان ہو جائے گی۔ الفاظ کا سن اور قرن متعین ہو سکے گا۔

کمپیوٹر میں کسی ایک عہد کے متون جس قدر زیادہ ہوں گے، ''اجنبیت'' اور ''فاصلے'' کا مطالعہ اتنا زیادہ واضح اور صحیح ہوگا۔ یوں اُردو میں متنی تنقید کو بھی سائنٹیفک تحقیق کا درجہ حاصل ہوگا۔ تاہم کمپیوٹر کسی بھی نسخے کو اساس قرار نہیں دے سکتا۔ یہ علمی دقتِ نظری کا کام ہے۔ بنیادی اصول یہ ہے کہ حروف، الفاظ یا متون کے شجرے کسی مواد کی مختلف اطراف کو منظم کرنے میں مدد دیتے ہیں اور ادبی کاموں کو سائنسی تحقیق کی بنیادیں فراہم ہوتی ہیں۔ مختلف متون کے درمیان نظریۂ اطلاعات کے استعمال سے علم زبان کے بہت سے شعبوں میں کام کی یکسانیت اور استناد حاصل ہوتا ہے۔ تجزیوں، درجہ بندیوں کے مسودات کے ماخذوں کا صحت کے ساتھ تعین ہو سکتا ہے۔ اس سے لغات اور الفاظ کو اطلاعیات، مواصلات اور موبائل کی دنیا میں زیادہ عمدہ اور بہتر انداز سے استعمال کرنے میں مدد ملے گی۔ اُردو میں تحقیق کی سمت بدل جائے گی اور ادب کی تخلیق میں بہتری، تفحص اور تعمق پیدا

ہوگا۔متون زیادہ عملی، منظّم اور واضح ہوسکیں گے۔تاریخ، قانون اور دستاویزی دنیا میں بھی ایک تکنیکی انقلاب برپا ہوگا۔ اُردو تحقیق کی بنیادیں زیادہ پختہ، گہری اور استوار ہوں گی۔(۲۰۰۸ء میں "دریافت"، اسلام آباد، "اخبار اُردو"، اسلام آباد، "الماس" خیر پور، "تحقیق" جام شورو اور "اورینٹل کالج میگزین" لاہور میں مختلف کی اشاعتوں کے بعد یکجا۔ جولائی ۲۰۰۸ء میں نظر ثانی کے بعد)

۱۰

اُردو اطلاعیات کی نصابی شمولیت

یونیسکو کی دستاویز برائے تیاری نصابِ اطلاعیات میں درج ہے کہ اطلاعیات (Informatics) ایک نیا میدانِ عمل ہے۔ نتیجے کے طور پر اطلاعیات کی تدریس مسلسل تبدیلیوں کا شکار رہتی ہے۔ وزارت تعلیم اور اعلیٰ تعلیمی کمیشنوں کو ان تبدیلیوں کے لیے یونیسکو کے تیار کردہ The Curriculum Framework, 2000, (ICF-2000) Informatics کو ملحوظ رکھنا چاہیئے۔ یہ دستاویز اس لیے تیار کی گئی ہے کہ ان تبدیلیوں سے عہدہ برآ ہونے کی اساس وضع کی جا سکے۔[1]

اطلاعیات کا یہ نصاب مقامی ضروریات کے مطابق مرتب ہونا چاہیئے۔ جہاں ابھی اطلاعیات (Informatics) ایک نیا میدان عمل ہے، وہیں اس کی ذیلی شاخیں بھی وضع ہونا شروع ہو چکی ہیں، جو اور بھی نئی ہیں۔ ان میں حیاتی اطلاعیات (Bio-Informatics) سے لے کر اُردو اطلاعیات (Urdu Informatics) تک ایک طویل فہرست سامنے آتی ہے۔

وکی پیڈیا میں "اطلاعیات" کی تعریف یہ کی گئی ہے کہ یہ اطلاعات و معلومات کی سائنس ہے جس میں اطلاعات کی عمل کاری اور اطلاعاتی نظام کی انجینئری کا احاطہ کیا جاتا ہے یعنی:

اطلاعات کی جمع آوری اور ترتیب اور تحصیل کا سارا عمل کمپیوٹر کی مدد سے انجام دینا "اطلاعیات" کہلاتا ہے۔ اس کے اصل الفاظ میں یہ تعریف مفصل طور پر یوں بیان ہوئی ہے۔[2]

> "**Informatics** is the science of information, the practice of information processing, and the engineering of information systems. Informatics studies the structure, algorithms, behavior, and interactions of natural and artificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields. Since the advent

of computers, individuals and organizations increasingly process information digitally. This has led to the study of informatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.

Used as a compound, in conjunction with the name of a discipline, as in medical informatics, bioinformatics, etc., it denotes the specialization of informatics to the management and processing of data, information and knowledge in the named discipline, and the incorporation of informatic concepts and theories to enrich the other discipline; it has a similar relationship to library science."

۱۹۹۸ء میں جب اُردو کی کمپیوٹر کاری (Computerization) کا عمل شروع ہوا تو کمپیوٹر پر اُردو کے کوائف (Data) کی عمل کاری (Processing) یا کوائف کاری (Data Processing) کے لیے "اُردو اطلاعیات" (Urdu Informatics) کی اصطلاح وجود میں آئی، جس میں اُردو کی حرف چینی/کتابت سے لے کر اُردو کی کمپیوٹری لسانیات اور اس کی انجینئری تک ایک وسیع میدان زیرِ عمل آیا۔ مقتدرہ قومی زبان کے مرکزِ فضیلت برائے اُردو اطلاعیات سے لے کر نیشنل یونیورسٹی فاسٹ لاہور کے CRULP تک کئی ادارے، مائیکروسافٹ کا مقامی کاری (Localization) پروگرام (LLP) اور موبائل فون کمپنیاں Nokia, Motorola وغیرہ اسی میدان کو مستحکم کرنے میں سرگرمِ عمل ہوئے۔ اُردو اطلاعیات کے بارے میں وکی پیڈیا میں یوں بیان ہوا ہے۔[۳]

Urdu Informatics (Urdu: اُردو اطلاعیات) relates to the cutting-edge research and efforts in bringing the utilities and usage of Urdu language to the modern information and communication technologies in education and businesses.

National Language Authority has been at the forefront in introducing Urdu Informatics as a tool for wider standardisation of the language. Apart from development of Urdu keyboard, one of the key steps in this respect has been the establishment of a Centre of Excellence for Urdu Informatics in Islamabad, Pakistan.Project Director Dr. Attash Durrani coined this term in 1998 to establish this discipline as a branch of Informatics۔"

ہائر ایجوکیشن کمیشن نے اپنے نصاب برائے اُردو (۲۰۰۳ء) میں ابھی ایم اے اور ایم فل کی سطح

کے "اُردو کمپیوٹر وانٹرنیٹ سے استفادہ" اور "لسانیات" کے کورس شامل کیے ہیں۔[۴]
قومی نصاب برائے اُردو کے لیے وفاقی وزارت تعلیم نے اُردو میں کمپیوٹر، انٹرنیٹ، ای میل وغیرہ کو، ایف اے تک نصاب ۲۰۰۶ء کا حصہ بنایا ہے۔[۵]
علامہ اقبال اوپن یونیورسٹی نے البتہ لسانی اطلاعیات کے اس پہلو کو اپنے کمپیوٹر سائنس کے شعبے میں شریک کیا ہے۔ BS اور MS کے تخصیصی میدانوں میں کچھ مضامین کا رخ اس سمت کیا ہے اور پی ایچ ڈی کے پروگرام میں فطری زبان کاری (Natural Language Processing) اور علمی انجینئری (Knowlege Engineering) کو اہمیت دی ہے، جو "اُردو اطلاعیات" کے لیے بھی بنیاد کا کام دیتی ہے۔
علم التعلیم (Education) کے میدان میں بھی علامہ اقبال اوپن یونیورسٹی کے کورس "تدریسیاتِ اُردو" (۲۰۰۳ء) برائے ماسٹر آف ایجوکیشن میں "تدریسِ اُردو اور کمپیوٹر" کے نام سے اور Text Book Development کورس (۲۰۰۶ء) میں "ای ٹیکسٹ بک" کے نام سے اجزاء شامل کیے گئے ہیں۔ لائبریری سائنس کے کورس میں "مسل داری" میں KWIC اور KWOC انداز میں اُردو مسل داری، اشاریہ سازی اور "آٹومیشن" کو شامل کیا گیا ہے۔
اب سوال یہ نہیں کہ "اُردو اطلاعیات" کا مضمون ہونا چاہیئے یا نہیں۔ سوال صرف یہ ہے کہ اسے کس طرح سے ہونا چاہیئے؟ بنیادی طور پر یہ میدان دو علوم سے تعلق رکھتا ہے۔ "اُردو" اور "اطلاعیات"۔ اُردو کے میدان میں اس کی ذیلی صورت "لسانیات" اور "قواعد" ہے اور مسئلہ یہ ہے کہ اُردو کا میدان ابھی ان میں بھی مہارت اور تخصص کی سطح تک نہیں پہنچا۔ "اطلاعیات" کا میدان بھی ابھی "کمپیوٹر سائنس" اور "انفارمیشن سائنس" کے مضامین ہی کے اندر غوطے کھا رہا ہے، جبکہ اطلاعیات محض کمپیوٹر کی اطلاقی صورت کا نام ہے۔
پہلے ہم "اُردو" کے میدان کو لیں۔ اس کی اعلیٰ نصابی سطح بھی صرف "ادبیات" ہی کا احاطہ کرتی ہے۔ زبان، لسانیات، ترجمہ کاری، اصطلاحیات، تدریسیات، اصولِ تحقیق اور ادب کے حوالے سے جدید ادبیات، پاکستانی ادبیات اور عالمی تقابلی ادبیات کا احاطہ بھی نہیں ہو رہا۔ "اُردو اطلاعیات" کے میدان میں ان تمام امور کی اہمیت ہے۔ ہائر ایجوکیشن کمیشن کے محولہ بالا جدید نصاب میں بھی یہ کہا گیا ہے:۔[۵]

"زبان کی حیثیت مسلمہ ہے، تاہم طلبہ میں ادب کے تصورات
کو زیادہ واضح انداز میں اجاگر کرنے کی ضرورت ہے۔"

یہ بیان نہ صرف زبان کی اہمیت کو کم بلکہ رد کرتا ہے۔ اُردو لازمی بی اے کے مضمون میں "قواعد" کے نام سے جو باب شامل ہے، اس میں بھی صرف ادبی اصطلاحات کا احاطہ کیا گیا ہے۔ زبان،

لسانیات، ترجمہ، اصطلاحیات، اطلاعیات کا تعارف اسی سطح پر ہو جانا چاہیئے تھا۔ اُردو لفظ سازی، اصولِ تحریر، پاکستانی اُردو، پاکستانی ادب اسی سطح کی چیزیں ہیں۔

ایم اے اُردو کے نصاب میں بھی تاریخِ ادبِ اُردو کے سیاسی پس منظر، نفسیاتی تاریخ، اُردو زبان کا ارتقاء، اُردو اطلاعیات کی بنیادی ضرورتوں، صرفیات (Morphology)، نحو (Syntax) وغیرہ کو تدریس میں شامل ہونا چاہیئے۔ تدوین کے بنیادی اصول، متنی انتقالات وغیرہ بھی کمپیوٹری استعمال کے ساتھ آنا چاہئیں۔

ایم فل اُردو کی سطح پر تو ایسے کمپیوٹری لسانیات (Computational Linguistics) کی تفاصیل اور اُردو ترجمہ کاری اور اصطلاحیات جیسے مضامین کے ساتھ شامل کرنا چاہیئے۔ تحقیق اور تدوین میں بھی کمپیوٹری استعمالات اس کا حصہ ہونا چاہئیں۔ یہ صورتیں ہوں گی تو اُردو کا ایم فل پاس طالبعلم اُردو اطلاعیات کے ایم فل کی بنیاد بن سکے گا۔

جہاں تک "اُردو اطلاعیات" کے اعلیٰ سطحی نصاب کا تعلق ہے، فی الوقت اس کے نفاذ کے امکانات تین اداروں میں ہیں:

۱۔ مرکزِ فضیلت برائے اُردو اطلاعیات، اسلام آباد

۲۔ وفاقی اُردو یونیورسٹی، اسلام آباد/کراچی

۳۔ نیشنل یونیورسٹی فاسٹ، لاہور

لیکن نچلی سطح پر یہ نصاب بہت سے اداروں میں شامل ہو سکتا ہے، مثلاً ماسٹر سطح پر یہ MCS اور MIT کے تخصیصی میدان کے طور پر MT اور سافٹ ویئر انجینئرنگ میں تخصص کے طور پر، بی ایس سی میں بعض حصوں میں ایریا یا تخصص کے انداز میں اسے استعمال میں لایا جا سکتا ہے۔

اُردو اطلاعیات ہی اُردو کا مستقبل ہے اور خود یہ مستقبل کے پیشہ ور اور تحقیق کار تیار کرنے کا ایک ایسا عمل ہے جو اپنے روزگار کے میدان خود پیدا کرے گا۔ ضرورتیں بدلتی رہیں گی اور کل کی نصابی ضروریات بھی مختلف ہوتی چلی جائیں گی۔ کل کی ضرورت کا کل کے ماہرین جائزہ لیں گے۔ فی الوقت اس کی تین سطحیں یوں ہوں گی:۔

۱۔ غیر پیشہ ورانہ سطح: اس میں کمپیوٹر کے میدان سے دور افراد کی تربیت کا عمل شامل ہے۔ انھیں اطلاعیات کا استعمال سکھانا مقصود ہے مثلاً اُردو کے طلبہ، تحقیق کار اور اساتذہ کو اس میں عمل کار (Processor) اور ابلاغیاتی نظام (Communication System)، تحقیق و تدوین کے آلات (Research and Editing Tools) وغیرہ کی تربیت دی جائے گی۔

۲۔ تکنیکی سطح: اُردو اطلاعیات کے تکنیک کار۔ مثلاً اُردو کے سسٹم انجینئر، اُردو سافٹ ویئر ڈویلپرز وغیرہ تربیت

پا ئیں گے۔

۳۔ پیشہ ورانہ سطح: اُردو اطلاعیات کے پیشہ ور جو فانٹ، مشینی ترجمہ، کوائفیہ، کوائف گھر اور دوسرے ایسے کاموں کو وضع کریں گے۔ یہ مختلف اداروں میں ان کی پیشہ ور مہارتوں کے لیے درکار ہوں گی۔

۴۔ ماہرانہ سطح: اُردو اطلاعیات کے لسانی اور انجینئرَی کے میدانوں میں مہارت اور تخصص کے حامل افراد جیسے پروفیسر، مینجر وغیرہ کی تربیت۔

ان سطحوں پر تمام افراد کے لیے عام تربیتی کورسوں کے علاوہ گریجویٹ اور پوسٹ گریجویٹ کورس درکار ہوں گے۔ تربیتی سطح کے لیے ۲۰ کریڈٹ پوائنٹ تک کے کورس کافی ہوں گے اور بنیادی آلات اور تصورات کی تعلیم بہت ہوگی۔ گریجویٹ سطح پر ۴۰ سے ۸۰ کریڈٹ پوائنٹ اور پوسٹ گریجویٹ پر ۱۶۰ کریڈٹ پوائنٹ بہت ہو سکتے ہیں۔

تربیتی سطح پر اطلاعیات کے آلات، سافٹ ویئر، اطلاعیات کی پیش کش اور تحصیل کے کورس ہوں گے۔ تکنیکی سطح پر الگورزم، سسٹم ڈیزائن، سافٹ ویئر کی تنصیب، کمپیوٹر سسٹم، آرکی ٹیکچر، سماجی اور اخلاقی اطلاقات وغیرہ جیسی مہارتیں درکار ہوں گی۔ پیشہ ورانہ سطح پر گریجویٹ ہارڈ ویئر اور سافٹ ویئر کے علاوہ کمپیوٹری لسانیات، کمپیوٹری کتابت وغیرہ کو ملحوظ رکھیں گے۔ تحقیق اور انجینئرَی اس کا حصہ ہوں گے۔ انھیں غیر اطلاعیاتی مضامین بھی پڑھنا ہوں گے جیسے "سائبر جرائم"، "اخلاقیات"، "نظم ونسق"، "قانون"، "نفسیات"، "لسانیات"، "فلسفہ" اور "سماجی علوم" کے متعلقہ پہلو۔ ماہرانہ سطح پر ہر قسم اور ہر پہلو پر تحقیق اور تکنیک شامل ہوگی۔

اطلاعیات کے مجوزہ نصاب میں یونیسکو کے مندرجہ ذیل خیالیے (Themes) شامل کیے گئے ہیں:-[۶]

1. Representation of Information
2. Formalising in Information Process
3. Information Modelling
4. Algorithmics
5. System Design
6. Software Development
7. Potentials of other Technologies
8. Computer Systems and Architecture
9. Computer Based Communication

10. Social and Ethical Informations

11. Personal and Interpersonal Skills.

12. Context of other Disciplines

زبان اور ٹیکنالوجی کے موضوع پر جو کانفرنسیں ہوئی ہیں، ان میں CLT-07 باڑہ گلی، اگست ۲۰۰۷ء قابل ذکر ہے جس میں مندرجہ ذیل موضوعات پیش نظر رکھے گئے۔۵؎ یہی موضوعات CLT-09 جنوری ۲۰۰۹ء نیشنل یونیورسٹی فاسٹ لاہور کے لیے بھی سامنے آئے ہیں۔ مجموعی طور پر یہ موضوعات حسب ذیل ہیں:

Anaphora, Semantics,Phonetics, Phonology, Morphology, Syntax, Corpus Linguistics, Discourse, Analysis, Wirting System, Ontologies, Text Summarization, Linguistic Resources, Speech Recognition, Text to Speech Analysis, OCR, Handwriting Recognition, Localization, Stemming, POS Tagging, Grammar Modelling, Fonts, Machine Translation, Information Retrieval and Computer Assisted Language Learning.

اگر ہم اس میں اُردو کے مندرجہ ذیل پہلو شامل کریں تو اُردو اطلاعیات کے نصاب کی بنیادیں استوار ہوتی نظر آتی ہیں:

۱۔اُردو لسانیات (صرفیات،صوتیات وغیرہ)

۲۔صرف ونحو

۳۔حروف اور کتابت

۴۔علمِ ترجمہ اور اصطلاحیات

۵۔علمِ تدوین اور تحقیق

پیشہ ورانہ سطح پر اس نصاب کا احاطہ مندرجہ ذیل امور تک وسیع ہوگا:

1.Font and OCR Development/ Script and Character Sets (حروف اور فانٹ)

2. Standardization and its Institutes/ ISO, UNICODE etc (معیار بندی).

3. Localization and Translating tools(مقامی ترجمہ کاری)

4.Speech to Text to Speech(آواز تا متن تا آواز)

5. Machine Translation/ Translatology(مشینی ترجمہ)

6. Data Base and Data Bank/ House (کوائف گھر)

7. Information Processing and Retrieval(اطلاعات کاری و بازگیری)

8. Natural Language Processing(فطری زبان کاری)

9. Knowledge Engineering and Thesaurus Building
(علمی انجینئری تھیسارس)

10. Document Engineering and Researching
(دستاویزی انجینئری و تحقیق کاری)

11. Pedagogy and Andragogy (E-Teaching-Learning)/ Fluid
Text and Tools(تدریسیات اطفال و بالغاں)

12.Web and Internet(ویب، انٹرنیٹ)

13. Informatics Business Management (E-Business etc.)
(برقیاتی تجارت)

14. E-Governance (Language Aspect)(برقیاتی حکومت)

15. Language Engineering/ Terminology Science
(لسانی انجینئری و اصطلاحیات)

"اُردو اطلاعیات" کے اس حیطہ ءکار کو جب ہم مختلف سطحوں پر نافذ کرتے ہیں تو ہمیں نظر آتا ہے کہ اعلیٰ ترین سطح پر اس کے لیے ڈاکٹریٹ اور پوسٹ ڈاکٹریٹ تعلیم کا انتظام ہونا چاہیئے۔ دنیا کی کئی یونیورسٹیوں میں تدریس اور تحقیق کے ایسے انتظامات موجود ہیں مثلاً Localization, Terminology Science, Translatology, Computational Linguistics, Documentation Engineering وغیرہ۔ ان تمام تجربات کے باہمی تبادل سے ایک ایسے ادارے یا انسٹی ٹیوٹ کا قیام عمل میں آ سکتا ہے جہاں "اُردو اطلاعیات" کو ایک اعلیٰ سطحی مضمون کی حیثیت دی جا سکے۔

اس سطح کے لیے تدریسیاتی طور پر ایم فل یا ایم ایس سطح کے فارغ التحصیل درکار ہوں گے، جنھیں تکنیکی و تحقیقی انداز میں تعلیم دی گئی ہو۔ اس سطح کے فارغ التحصیل طلبہ دیگر اداروں میں پیشہ ورانہ، تکنیکی اور تدریسی خدمات بھی انجام دینے کے لیے تیار ہو سکتے ہیں۔

ایم فل یا ایم ایس اُردو اطلاعیات کے لیے طلبہ دو اطراف سے آئیں گے۔ پہلی سمت سے تو کمپیوٹر

سائنس کے ایم فل یا ایم ایس ہوں جنھیں یہاں اُردو کی کمپیوٹری لسانیات اور سافٹ ویئر کی تیاری وغیرہ کی تعلیم دی جائے گی اور دوسری سمت اُردو کے ایم فل گریجویٹ کی ہے جنھیں یہاں سافٹ ویئر انجینئری، لوکلائزیشن وغیرہ کی تعلیم دی جائے گی اور پھر ایم فل/ ایم ایس اُردو اطلاعیات کی تخصیص کی اجازت دی جائے گی۔

ہر دو سمتوں کے لیے الگ الگ بنیادی کورس درکار ہوں گے۔ مثلاً کمپیوٹر سائنس کے گریجویٹ اور پوسٹ گریجویٹ سطح پر اطلاعیات اور کمپیوٹری لسانیات وغیرہ پر توجہ دینا ضروری ہوگا اور اُردو کے میدان میں بھی اطلاعیات کا تعارف اور کمپیوٹری استعمالات وغیرہ کی تعلیم درکار ہوگی۔

مرکزِ فضیلت برائے اُردو اطلاعیات کے کارکنوں نے "اُردو اطلاعیات" کی نصاب سازی کے چند رہنما اصول وضع کیے ہیں جو حسب ذیل ہیں: ے

0. Prerequisits

0.1. Training Tutorials

0.2. Undergraduates

0.3. For Urdu Graduates

0.4. For Computer Science/ IT Graduates

0.5. MS/M Phil. Urdu Informatics

0.6. Ph.D./ D.U.I. Urdu Informatics

ان سطحوں کے کورسوں کے نام حسب ذیل ہو سکتے ہیں:

0. <u>Prerequisits</u>

0.1. Urdu Informatics- An Introduction :اُردو اطلاعیات ۔۔ایک تعارف

0.2. Cyber Ethics/ Crimes.:سائبر اخلاقیات/ جرائم

0.3. Localization Business/ Industry:مقامی کاری تجارت/صنعت

0.4. History and Working of Computer :کمپیوٹر: تاریخ وعمل

0.5. Computer in Education :تعلیم میں کمپیوٹر

0.6. Sociology of I.T. :اطلاعیاتی سماجیات

0.7. Computer Management :کمپیوٹری انتظام

0.8. Computer Application in Urdu :اُردو میں کمپیوٹری آلات واستعمال

0.9. Needs of Software Engineering (Linguistics Aspects)

سافٹ ویئر انجینئری کی ضروریات (لسانی پہلو):

1.<u>Training Tutorials:</u>

1.1. <u>Hardware & Networking:</u>

(Microprogramming, Logic, Memory, Data,

Implementation, Circuits, Performance,
Technology etc.)

1.2. Logical Tools:
(Transfer, Urdu Windows, Vista, Office etc. other search and research tools)

1.3. Software:
(Fonts, M.T., D.B. etc.)

2. Undergraduates:

2.1. Software Engineering:
(Lingual requistions and needs, Architecture, Testing, Programming, Operationg System)

2.2. Localizaion and Standardization:
(Local Culture and Languages, Globalization, Standard Tools, Standard Institutes= ISO, Unicode etc.)

2.3. Data Development and Retrieval
(Data Structure, Storages, Encryption, Coding, Information Theory, Filing, Retrievals etc.)

2.4. Mathematics for Computing
(Algebra, Algorithm, Graphics etc.)

2.5. Lingual Information Systems
(General to Specific, Database, Storage, Databanks/ Houses, Applications, Interfaces, Sounds etc.)

2.6. Natural Language & Processing Methodologies
(A.I., Graphics, Images, Patterns, Simulation, Modelling, Documentation, Researching, e-publishing etc.)

2.7. Font/OCR/Calligraphy
(Standards, Normalization, Secracy, Proofing Tools, Implementation etc.)

2.8. Machine Translation Tools
(Urdu Target Language in Example Base and Knowledge Base Applications)

3. Urdu Graduates

3.1.

3.2. مذکورہ مجوزہ فہرستوں میں سے

3.3.

4. Computer/ IT Graduates

4.1.

4.2. مذکورہ مجوزہ فہرستوں میں سے

4.3.

5. M.S/ M. Phil- Urdu Informatics

5.1.

5,2. مذکورہ مجوزہ فہرستوں میں سے

5.3.

6. Ph.D./ D.U.I. Urdu Informatics

اُردو اطلاعیات کا تدریسیاتی اور انتظامی پہلو بھی قابل غور ہے۔ تدریسیاتی حوالے سے ایک پہلو خود کمپیوٹر کی تعلیم اُردو میں دینے کا ہے جسے اصولی طور پر وفاقی اُردو یونیورسٹی سے شروع ہونا چاہیئے تھا۔ مگر وہاں اور سب کچھ چلتا ہے، صرف اُردو میں تدریس نہیں ہوتی۔ تاہم امید افزا صورت یہ ہے کہ مائیکروسافٹ نے IBA کراچی کے ڈاکٹر حبیب خان کی مدد سے کم از کم لوکلائزیشن کے پہلو سے ایسے اُردو نصاب کی تیاری کا اہتمام کیا ہے اور اُردو میں کمپیوٹر کی تدریس کا یہ نصاب وضع کیا جا رہا ہے۔ مرکزِ فضیلت نے اُردو ونڈوز، آفس، وسٹا، اُردو فانٹ، ترجمہ کاری اور دوسرے ایسے امور کی تربیت کے لیے مواد، ٹیوٹوریل، اسباق تیار کیے اور عملاً تربیت کا اہتمام کیا ہے، جنھیں مستقبل کے لیے بنیاد ٹھہرنا ہے۔

ان کورسوں کے کریڈٹ آور، ترتیب اور ترجیح، تقدیم اور تاخیر کا انتظام متعلقہ ادارہ، انسٹی ٹیوٹ، یونیورسٹی وغیرہ دیگر تفاصیل کے تحت قائم کر سکے گا۔ ہائر ایجوکیشن کمیشن ان نکات کی روشنی میں نصاب کمیٹی تشکیل دے سکتا ہے۔ اس میں کمپیوٹر سائنس، انفارمیشن سائنس، انفارمیٹکس، اُردو اور علم التعلیم و نصاب سازی سے متعلقہ اراکین شامل ہوں۔ یہ سب باہم غور و فکر کے بعد ہی اُردو اطلاعیات کا ہر سطح کا نصاب وجود میں لا سکیں گے۔

"اُردو اطلاعیات" کے لیے مرکزِ فضیلت برائے اُردو اطلاعیات کا مجوزہ انسٹی ٹیوٹ ان بنیادوں پر آگے بڑھ رہا ہے کہ اس میں R & D یعنی تحقیق و ترقی کا بنیادی کام انجام دیا جائے۔ اس کے ساتھ ساتھ "اُردو" اور "کمپیوٹر سائنس" کے نصاب (گریجویٹ اور پوسٹ گریجویٹ) سطح پر اُردو اطلاعیات کے اساسی تصورات، عملی تربیت اور رجحانات پیدا کیے جائیں۔ علامہ اقبال اوپن یونیورسٹی اسلام آباد نے بی سی ایس سطح پر ایسے کورسوں کا اہتمام کیا ہے، اسی طرح پی ایچ ڈی تک بعض کورس شامل کیے ہیں۔ پشاور یونیورسٹی کے شعبہ کمپیوٹر سائنس میں مشینی ترجمے کی تخصیص کی جا رہی ہے۔ ایم اے اور پی ایچ ڈی اُردو کے نصاب میں تو "کمپیوٹر کی تاریخ"، "عمل"، "اُردو میں کمپیوٹری آلات و استعمال" اور "تعلیم میں کمپیوٹر" جیسے

کورسوں کا نام دینا ہوگا۔ "کمپیوٹر سائنس" کے ایم سی ایس میں "مقامی کاری"، "سافٹ ویئر انجینئری کی ضروریات"، "لسانی انجینئری" جیسے کورس شامل کرنا ہوں گے۔

چونکہ اطلاعیات لائبریری سائنس کی بنیادیں بھی رکھتا ہے، اس لیے موجودہ ایم ایل ایس سی میں بھی ایسے کورس شامل کرنا ہوں گے جو اُردو میں لائبریری سائنس اور انفارمیشن سائنس کو فروغ دینے کا سبب بن سکیں۔

جہاں تک تدریسی طریقوں کا تعلق ہے، فرسودہ لیکچر، نصابی کتابیں، نوٹس اور امتحانات کام نہیں دیں گے۔ بنیادی تجربہ گاہ اور عملی، متعامل (Interactive) طریقِ تدریس و تربیت درکار ہوں گے۔ لسانی تجربہ گاہ (Language Lab) کا استعمال ہوگا۔ اعلیٰ آلات تحقیق اور فائبر لنک کی سہولتیں درکار ہوں گی۔ اس حوالے سے جہاں "اُردو اطلاعیات" کا مضمون سائنسی و تکنیکی بنیادوں پر استوار ہوگا۔ وہیں "اُردو" اور "لائبریری سائنس" کے مضامین میں شامل کیے جانے والے اطلاعیاتی و لسانیاتی کورسوں کے لیے ان آلات اور تجربہ گاہوں کی ضروریات بھی ہوں گی۔ یوں یہ مضامین بھی "سائنسی" قرار پائیں گے۔ لسانیات کا کورس تو کسی بھی طرح "لسانی تجربہ گاہ" کے بغیر اپنے نتائج نہیں دے سکتا۔ پھر کم از کم ایم فل اور پی ایچ ڈی کی سطح پر اُردو کی لسانی پہلو سے تدریس کیوں کر انجام پا سکتی ہے۔ اُردو کے اساتذہ کے لیے نصاب پر ایک بار پھر نظر واپسیں ڈالنے کا مشورہ ہے۔ اُردو کے نصاب میں "اطلاعیات" جگہ پائے گی تو "اُردو اطلاعیات" کا ڈسپلن پروان چڑھ سکے گا۔

حوالہ جات:

1. UNESCO, **The Informatics Curriculum Framework, 2000, for Higher Education**, Paris, 2000, P-iii.

2. Wikipedia.... **Informatics**

3. Wikipedia... **Urdu Informatics**

4. HEC, **Curriculum of Urdu for B.A./ M.A./ M. Phil**., 2003, PP-19,29,31.

5. Department of Computer Science, **Proceedings of the Conference on Language and Technology** (CLT-07), Bara Gali, University of Peshawar, 7-11 August 2007.

6. Ministry of Education, **National Curriculum for Urdu**, Islamabad, 2006.

7. UNESCO, **op.cit,** PP.31-33

8. CEUI, **An Outline for Curriculum Proposals for Urdu Informatics** (Develped by: Mahboob Khan, Ghulam Rasool, Abdul Quddus Khan, Zeeshan Ranjha, Quratul-Ain), Islamabad, July 2008.

9.www.nlauit.gov.pk/tutorials (Prepared by: Akbar Sajjad, Sana Nisar)

ویب سائٹیں:

1. www.nlauit.gov.pk
2. dcs.edu.pk
3. Unesco.org/ICF-2000
4. ifip.or.at
5. microsoft.com
6. crulp.org
7. uchicago.edu
8. wikipedia.org

(شعبہ اُردو پشاور یونیورسٹی اور ہائر ایجوکیشن کمیشن کی مشترکہ عالمی اُردو کانفرنس "اعلیٰ تعلیم: تحقیقی و تدریسی مسائل اور ان کا حل" ۷ تا ۱۱۔ اگست ۲۰۰۸ء، باڑہ گلی میں پڑھا گیا)۔

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *YES*
   If YES explain *This is a significant revision of earlier proposals submitted to UTC. See references above. This is a revision of a document submitted on April 18, 2008, L2/08-159.*
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? *YES*
   If YES, with whom? *National Language Authority, Pakistan*
   If YES, available relevant documents: *See references above.*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *NO*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare) *Common*
   Reference:
5. Are the proposed characters in current use by the user community? *YES*
   If YES, where? Reference: *Pedagogical materials*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *YES*
   If YES, is a rationale provided? *NO*
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *YES*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *NO*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *NO*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? *NO*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *NO*
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? *NO*
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *NO*
    If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility character(s)? *NO*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
    If YES, reference:

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. **Title:** ***Proposal to encode 16 characters for Arabic pedagogical use***
2. Requester's name: *Dr. Attash Durrani (NLA, Pakistan) ; Kamal Mansour ; Rick McGowan*
3. Requester type (Member body/Liaison/Individual contribution): *Member body*
4. Submission date: *June 24, 2008*
5. Requester's reference (if applicable): *L2/06-039 L2/06-240 L2/07-094 L2/07-174*
6. Choose one of the following:
   This is a complete proposal: *YES*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *YES*
      Name of the existing block: *(Arabic)*
2. Number of characters in proposal: *16*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary *A* B.1-Specialized (small collection) ____ B.2-Specialized (large collection) ____
   C-Major extinct ____ D-Attested extinct ____ E-Minor extinct ____
   F-Archaic Hieroglyphic or Ideographic ____ G-Obscure or questionable usage symbols ____
4. Is a repertoire including character names provided? *YES*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *YES*
   b. Are the character shapes attached in a legible form suitable for review? *YES*
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? *Kamal Mansour (Monotype Imaging)*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: *Truetype font – used to create this proposal – Kamal Mansour*
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *See L2/06-039*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *See L2/06-039, and othe references above*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *NO*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[1] Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03)

0880 Arabic Pedagogical Symbols 089F

	x88	x89
0	xx80	
1	xx81	
2	xx82	
3	xx83	xx93
4	xx84	xx94
5	xx85	xx95
6	xx86	
7	xx87	
8	xx88	
9	xx89	
A		
B	xx8B	
C	xx8C	
D	xx8D	
E		
F		

0880 ARABIC SINGLE NUQTA ABOVE
0881 ARABIC SINGLE NUQTA BELOW
0882 ARABIC DOUBLE NUQTA ABOVE
0883 ARABIC DOUBLE NUQTA BELOW
0884 ARABIC TRIPLE NUQTA ABOVE
0885 ARABIC TRIPLE NUQTA BELOW
0886 ARABIC TRIPLE INVERTED NUQTA ABOVE
0887 ARABIC TRIPLE INVERTED NUQTA BELOW
0888 ARABIC QUADRUPLE NUQTA ABOVE
• Sindhi
0889 ARABIC QUADRUPLE NUQTA BELOW
• Sindhi
088A <reserved>
088B ARABIC DOUBLE DANDA BELOW
• Sindhi
088C ARABIC DOUBLE NUQTA VERTICAL ABOVE
• Sindhi
088D ARABIC DOUBLE NUQTA VERTICAL BELOW
• Sindhi
088E <reserved>
088F <reserved>
0890 <reserved>
0891 <reserved>
0892 <reserved>
0893 ARABIC SINGLE CIRCLE BELOW
•Pashto
0894 ARABIC TOTA ABOVE
• Urdu
0895 ARABIC TOTA BELOW
• Urdu

Printed using UniBook™ (http://www.unicode.org/unibook/) *Date: 24-Jun-2008* *1*

Appendix-1

JTC1/SC2/WG2 N3460

L2/08-159R

Doc Type: Working Group Document
Title: Proposal to Encode 16 Characters for Arabic Pedagogical Use
Source: Attash Durrani (Pakistan NLA), Kamal Mansour, Rick McGowan
Status: Individual Contribution
Action: For National Body Review and Feedback
Date: June 24, 2008

References: L2/06-039 L2/06-240 L2/07-094 L2/07-174
http://www.unicode.org/L2/L2006/06039-urdu-arabic-prelim.pdf
http://www.unicode.org/L2/L2006/06240-durrani.pdf
http://www.unicode.org/L2/L2007/07094-regarding-nuqta.pdf
http://www.unicode.org/L2/L2007/07174-nuqta-disc.pdf

There is a need to encode dot patterns used in Arabic letters for pedagogical and other purposes, as exemplified in the original proposal L2/06-039 and related documents.

That proposal requested 22 characters to be encoded within the 0600 block, but there is a better open area of right-to-left codepoints tentatively reserved for Arabic additions in the range from 0880 through 08BF. This range is probably the best place to put a set of spacing dot pattern characters.

At meeting #115, UTC accepted 16 characters for encoding based on a previous version of this document. Characters not accepted have been removed from this revision.

A good deal of required information and examples are contained in the previous proposal L2/06-039, as well as discussion documents L2/06-240, L2/07-094 and L2/07-174. That information will not be re-iterated here.

The proposed 16 characters, their glyphs, names, and annotations are shown on the accompanying chart and nameslist.

These would be given characters properties as follows (in the syntax of the UnicodeData.txt file):

```
08nn;---NAME---;Sk;0;AL;;;;;N;;;;;
```

The line-breaking behavior of these characters would be similar to ordinary symbols of the "Sk" class.

Bibliography:

1. Muqaalaat-e-Hafiz Mehmood Sheerani, No.VIII, (ed. Mazhar Mahmood Sheerani), Published by Majlis Tarraqi-i-Adab, Lahore.

2. The Standard Urdu-English Dictionary by BaBa-e-Urdu Molvi Abdul-Haq, published by Anjman-e-Traqqi-e-Urdu, 1985.

3. Farhang-e-Talaffuz by Shaan-ul-Haq Haqqi published by National Language Authority, Pakistan, 1995.

4. Sheena Qaida, publised by N.W.F.P Government, Pakistan (2004).

5. Balti Qaida, published by National Language Authority of Pakistan, 2004.

6. Jonathan Kew, Encoding Arabic extensions: options for the future of Unicode, Unicode: Document L2/03/004, February 11, 2003.

7. The Unicode Standard, (version) 5.0, The Unicode Consortium, published by Addison-Wesley, Upper Saddle River, NJ, October 2006.

8. Tanveer Fatima, Ghost Characters Theory, nlauit.gov.pk.

9. Urdu Informatics (Vol, II), (ed. Dr. Attash Durrani), published by National Language Authority (Pakistan), 2008.

10. nlauit.gov.pk/e.magazine

Double Combinations Unicode Examples
(Up/Below Nuqta Characters)

Double Combinations Unicode Examples (Upper Nuqta Characters)

Single Combination Chart
(Upper Nuqta Characters)

⁘	∵	∴	⁝	¨	·	ؕ	″	=	¯	°	^	ˇ	ؔ	ؑ	
ا⁘	ا∵	ا∴	ا⁝	ا¨	ا·	اؕ	ا″	ا=	ا¯	ا°	ا^	اˇ	اؔ	اؑ	ا
ٮ⁘	ٮ∵	ٮ∴	ٮ⁝	ٮ¨	ٮ·	ٮؕ	ٮ″	ٮ=	ٮ¯	ٮ°	ٮ^	ٮˇ	ٮؔ	ٮؑ	ٮ
ح⁘	ح∵	ح∴	ح⁝	ح¨	ح·	حؕ	ح″	ح=	ح¯	ح°	ح^	حˇ	حؔ	حؑ	ح
د⁘	د∵	د∴	د⁝	د¨	د·	دؕ	د″	د=	د¯	د°	د^	دˇ	دؔ	دؑ	د
ر⁘	ر∵	ر∴	ر⁝	ر¨	ر·	رؕ	ر″	ر=	ر¯	ر°	ر^	رˇ	رؔ	رؑ	ر
س⁘	س∵	س∴	س⁝	س¨	س·	سؕ	س″	س=	س¯	س°	س^	سˇ	سؔ	سؑ	س
ص⁘	ص∵	ص∴	ص⁝	ص¨	ص·	صؕ	ص″	ص=	ص¯	ص°	ص^	صˇ	صؔ	صؑ	ص
ط⁘	ط∵	ط∴	ط⁝	ط¨	ط·	-	ط″	ط=	ط¯	ط°	ط^	طˇ	طؔ	طؑ	ط
ع⁘	ع∵	ع∴	ع⁝	ع¨	ع·	عؕ	ع″	ع=	ع¯	ع°	ع^	عˇ	عؔ	عؑ	ع
ڡ⁘	ڡ∵	ڡ∴	ڡ⁝	ڡ¨	ڡ·	ڡؕ	ڡ″	ڡ=	ڡ¯	ڡ°	ڡ^	ڡˇ	ڡؔ	ڡؑ	ڡ
ٯ⁘	ٯ∵	ٯ∴	ٯ⁝	ٯ¨	ٯ·	ٯؕ	ٯ″	ٯ=	ٯ¯	ٯ°	ٯ^	ٯˇ	ٯؔ	ٯؑ	ٯ
ک⁘	ک∵	ک∴	ک⁝	ک¨	ک·	کؕ	ک″	ک=	ک¯	ک°	ک^	کˇ	کؔ	کؑ	ک
گ⁘	گ∵	گ∴	گ⁝	گ¨	گ·	گؕ	گ″	گ=	گ¯	گ°	گ^	گˇ	گؔ	گؑ	گ
ل⁘	ل∵	ل∴	ل⁝	ل¨	ل·	لؕ	ل″	ل=	ل¯	ل°	ل^	لˇ	لؔ	لؑ	ل
م⁘	م∵	م∴	م⁝	م¨	م·	مؕ	م″	م=	م¯	م°	م^	مˇ	مؔ	مؑ	م
ں⁘	ں∵	ں∴	ں⁝	ں¨	ں·	ںؕ	ں″	ں=	ں¯	ں°	ں^	ںˇ	ںؔ	ںؑ	ں
و⁘	و∵	و∴	و⁝	و¨	و·	وؕ	و″	و=	و¯	و°	و^	وˇ	وؔ	وؑ	و
ہ⁘	ہ∵	ہ∴	ہ⁝	ہ¨	ہ·	ہؕ	ہ″	ہ=	ہ¯	ہ°	ہ^	ہˇ	ہؔ	ہؑ	ہ
ء⁘	ء∵	ء∴	ء⁝	ء¨	ء·	ءؕ	ء″	ء=	ء¯	ء°	ء^	ءˇ	ءؔ	-	ء
ی⁘	ی∵	ی∴	ی⁝	ی¨	ی·	یؕ	ی″	ی=	ی¯	ی°	ی^	یˇ	یؔ	یؑ	ی
ھ⁘	ھ∵	ھ∴	ھ⁝	ھ¨	ھ·	ھؕ	ھ″	ھ=	ھ¯	ھ°	ھ^	ھˇ	ھؔ	ھؑ	ھ
ے⁘	ے∵	ے∴	ے⁝	ے¨	ے·	ےؕ	ے″	ے=	ے¯	ے°	ے^	ےˇ	ےؔ	ےؑ	ے

sequence of digits follows them in logical order.

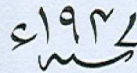
Poetic Verse Sign. U+060E ARABIC POETIC VERSE SIGN is a special symbol often used to mark the beginning of a poetic verse. Although it is similar to U+0602 ARABIC FOOTNOTE MARKER in appearance, the poetic sign is simply a symbol. In contrast, the footnote marker is a format control character that has complex rendering in conjunction with following digits. U+060F ARABIC SIGN MISRA is another symbol used in poetry.

Thus atomization -combination rules were already presented in the UNICODE, but without any theoretical elaboration that is now formed to achieve a good model for Arabic Script. This helps to develop a chart of atom combinations for character formation. Next are the Single Combination and Double Comibnation charts. There may be 660 letter formations of single combination depicted in next two pages and out of 20460 possible letter formation only those characters are mentioned which are already on Unicode.

One can easily find that the Ghost Charaters Theory saves the place on International Standards, i.e. (22+30) 52 spaces and 20460 possible letters are made to happen from past to far future of the Arabic Script to be used by any language of the world or for any pedogogical teaching purposes.

(Presented in the Shaikh Ayaz International Conference on Language and Literature (SAICLL), Sindh University, Jamshoro, on November 7-10 2008)

There is another example of Ghost Characters combination in the UNICODE i.e. ARABIC SIGN SANAH. It indicates a year (that is, as part of a date). This sign is rendered below the digits of the number it precedes. Its appearance is a vestigial form of the Arabic word for year, /sartatu/ (seen noon teh-marbuta), but it is now a sign in its own right and is widely used to mark a numeric year even in non-Arabic languages where the Arabic word would not be known. The use of the year sign is illustrated in Figure 8-6.

Figure 8-6. Arabic Year Sign

U+0602 ARABIC FOOTNOTE MARKER is another of these signs; it is used in the Arabic script in conjunction with the footnote number itself. It also precedes the digits in logical order and is written to extend underneath them.

Finally, U+0603 ARABIC SIGN SAFHA functions as a page sign, preceding and extending under a sequence of digits for a page number.

Like U+06DD ARABIC END OF AYAH, all of these signs can span multiple-digit numbers, rather than just a single digit. They are not formally considered combining marks in the sense used by the Unicode Standard, although they clearly interact graphically with the sequence of digits that follows them. They precede the sequence of digits that they span, rather than following a base character, as would be the case for a combining mark. Their General Category value is Cf (format control character). Unlike most other format control characters, however, they should be rendered with a visible glyph, even in circumstances where no suitable digit or

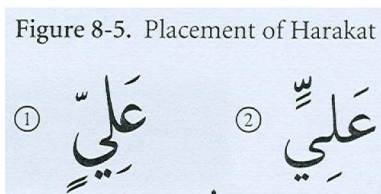
further discussion of joiners and non-joiners, see Section 16.2, Layout Controls.

Harakat are marks that indicate vowels or other modifications of consonant letters. The code charts depict a character in the harakat range in relation to a dashed circle, indicating that this character is intended to be applied via some process to the character that precedes it in the text stream (that is, the base character).

General rules for applying non-spacing marks are given in Section 7.9, Combining Marks of the Unicode (7:271). The few marks that are placed after (to the left of) the base character are treated as ordinary spacing characters in the Unicode Standard. The Unicode Standard does not specify a sequence order in case of multiple harakat applied to the same Arabic base character, as there is no possible ambiguity of interpretation. For more information about the canonical ordering of nonspacing marks, see Section 2.11, Combining Characters, and Section 3.11, Canonical Ordering Behavior.

The placement and rendering of vowel and other marks in Arabic strongly depends on the typographical environment or even the typographical style. Code Charts, the default position of U+0651 t) ARABIC SHADDA is with the glyph placed above the base character, whereas for U+064D ,) ARABIC KASRATAN the glyph is placed below the base character, as shown in the first example in Figure 8-5. However, computer fonts often follow an approach that originated in metal typesetting and combine the kasratan with shadda in a ligature placed above the text, as shown in the second example in Figure 8-5.

Figure 8-5. Placement of Harakat

① عَلِيٍّ ② عَلِيٍّ

harakat, which are applied to consonantal base letters. In normal writing, however, these harakat are omitted.

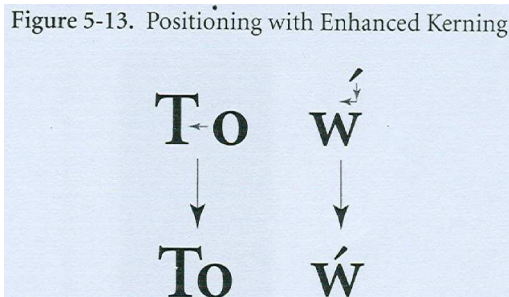
The Arabic script is written from right to left. Conformant implementations of Arabic script must use the Unicode Bidirectional Algorithm to reorder the memory representation for display Unicode Standards #9.

The Unicode Standard encodes the basic Arabic characters in the same relative positions as in ISO/IEC 8859-6. ISO/IEC 8859-6, in turn, is based on ECMA-114, which was based on ASMO 449.

Most punctuation marks used with the Arabic script are not given independent codes (that is, they are unified with Latin punctuation), except for the few cases where the mark has a significantly different appearance in Arabic-namely, U+060C ARABIC COMMA, U+061B ARABIC SEMICOLON, U+061E ARABIC TRIPLE DOT PUNCTUATION MARK, U+061F ARABIC QUESTION MARK, and U+066A ARABIC PERCENT SIGN. For paired punctuation such as parentheses, the glyphs chosen to represent U+OO28 LEFT PARENTHESIS and U+OO29 RIGHT PARENTHESIS will depend on the direction of the rendered text.

The Unicode Standard provides two user-selectable formatting codes: U+200C ZERO WIDTH NON-JOINER and U+200D ZERO WIDTH JOINER (see Figure 8-2, Figure 8-3, and Figure 8-4). The use of a non-joiner between two letters prevents those letters from forming a cursive connection with each other when rendered. Examples include the Persian plural suffix, some Persian proper names, and Ottoman Turkish vowels. The use of a joiner adjacent to a suitable letter permits that letter to form a cursive connection without a visible neighbor. This provides a simPle way to encode some special cases, such as exhibiting a connecting form in isolation. For

for positioning diacritics is an extension of the normal process of kerning to be both horizontal and vertical (see Figure 5-13). Typically, kerning maps from pairs of glyphs to a positioning offset. For example, in the word "To" the "0" should nest slightly under the "T". An extension of this system maps to both a vertical and a horizontal offset, allowing glyphs to be positioned arbitrarily.

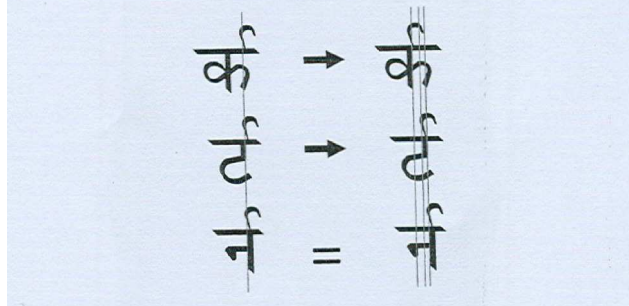
Figure 5-13. Positioning with Enhanced Kerning

For effective use in the general case, the kerning process must be extended to handle more than simple kerning pairs, as multiple diacritics may occur after a base letter.

Positioning with enhanced kerning can be combined with the ligature method so that in specific cases ligatures can be used to produce fine variations in position and shape.

In Arabic: U+0600-U+06FF we may fit this atomization combination. The Arabic script is used for writing the Arabic language and has been extended to represent a number of other languages, such as Persian, Urdu, Pashto, Sindhi, and Kurdish, as well as many African languages. Urdu is often written with the ornate Nastaliq script variety. Some languages, such as Indonesian/Malay, Turkish, and Ingush, formerly used the Arabic script but now employ the Latin or Cyrillic scripts.

The Arabic script is cursive, even in its printed form. As a result, the same letter may be written in different forms depending on how it joins with its neighbors. Vowels and various other marks may be written as combining marks called

positioning with enhanced kerning can be applied. If they are not available, then a fallback method can be used.

Figure 5-12. Positioning with Contextual Forms

Positioning with Contextual Forms. A more general method of dealing with positioning of nonspacing marks is to use contextual formation (see Figure 5-12). In this case for Devanagari, a consonant RA is rendered with a nonspacing glyph (reph) positioned above a base consonant. (See "Rendering Devanagari" in Section 9.1, Devanagari.) Depending on the position of the stem for the corresponding base consonant glyph, a contextual choice is made between reph glyphs with different side bearings, so that the tip of the reph will be placed correctly with respect to the base consonant's stem. Base glyphs generally fall into a fairly small number of classes, depending on their general shape and width, so a corresponding number of contextually distinct glyphs for the nonspacing mark suffice to produce correct rendering.

In general cases, a number of different heights of glyphs can be chosen to allow stacking of glyphs, at least for a few deep. This method can be combined with the ligature method so that in specific cases ligatures can be used to produce fine variations in position and shape.

Positioning with Enhanced Kerning. A third technique

Positioning with Ligatures. A fixed set of combining character sequences can be rendered effectively by means of fairly simple substitution (see Figure 5-11). Wherever the glyphs representing a sequence of <base character, non-spacing mark> occur, a glyph representing the combined form is substituted. Because the non-spacing mark has a zero advance width, the composed character sequence will automatically have the same width as the base character. More sophisticated text rendering systems may take additional measures to account for those cases where the composed character sequence kerns differently or has a slightly different advance width than the base character.

Positioning with ligatures is perhaps the simplest method of supporting non-spacing marks. Whenever there is a small, fixed set, such as those corresponding to the precomposed characters of ISO/IEC 8859-1 (Latin-1), this method is straightforward to apply. Because the composed character sequence almost always has the same width as the base character, rendering, measurement, and editing of these characters are much easier than for the general case ofligatures.

If a combining character sequence does not form a ligature, then either positioning with contextual forms or

Figure 5-11. Positioning with Ligatures

a + ◌̈ → ä

A + ◌̈ → Ä

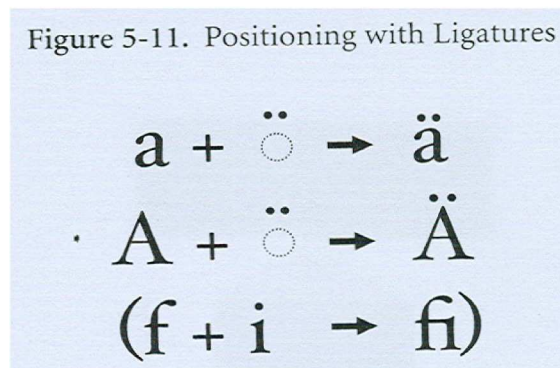
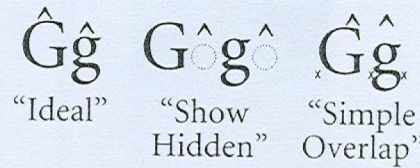
(f + i → ﬁ)

as an individual unit, with the non-spacing mark positioned on a dotted circle.

Figure 5-8. Fallback Rendering

Another method (Simple Overlap) uses a default fixed position for an overlapping zerowidth nonspacing mark. This position is generally high enough to make sure that the mark does not coUide with capital letters. This will mean that this mark is placed too high above many lowercase letters. For example, the default positioning of a circumflex can be above the ascent, which will place it above capital letters. Even though the result will not be particularly attractive for letters such as g-circumflex, the result should generally be recognizable in the case of single nonspacing marks.

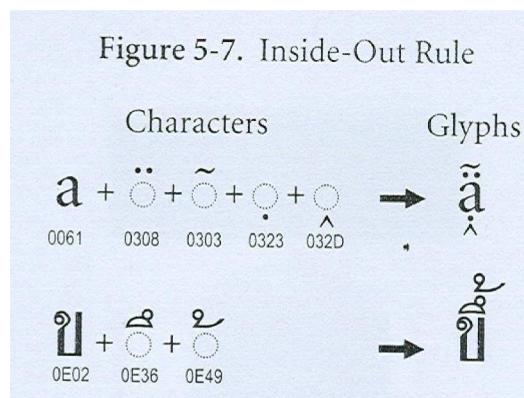
In a degenerate case, a nonspacing mark occurs as the first character in the text or is separated from its base character by a line separator, paragraph separator, or other format character that causes a positional separation. This result is caUed a defective combining character sequence (see Sectio/l 3.6, Combination). Defective combining character sequences should be rendered as if they had a no-break space as a base character.

When a quesion of Positioning Methods arose, Unicode resolved it as: (7:176-178):

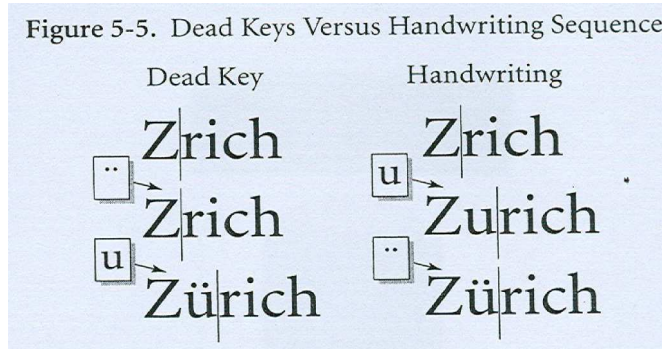
A number of methods are available to position non-spacing marks so that they are in the correct location relative to the base character and previous non-spacing marks.

beneath the base character, and the second nonspacing mark should appear below the first (see Section 2.11, Combining Characters). This default treatment of multiple, potentially interacting nonspacing marks is known as the inside-out rule (see Figure 5-7).

This default behavior may be altered based on typographic preferences or on knowledge of the specific orthographic treatment to be given to multiple nonspacing marks in the context of a particular writing system. For example, in the modern Vietnamese writing system, an acute or grave accent (serving as a tone mark) may be positioned slightly to one side of a circumflex accent rather than directly above it. If the text to be displayed is known to employ a different typographic convention (either implicitly through knowledge of the language of the text or explicitly through rich text-style bindings), then an alternative positioning may be given to multiple non-spacing marks instead of that specified by the default inside-out rule.

Figure 5-7. Inside-Out Rule

Several methods are available to deal with an unknown composed character sequence that is outside of a fixed, renderable set (see Figure 5-8). One method (Show Hidden) indicates the inability to draw the sequence by drawing the base character first and then rendering the non-spacing mark

non-spacing marks (see Figure 5-5).

Figure 5-5. Dead Keys Versus Handwriting Sequence

In the case of handwriting sequence, each keystroke produces a distinct, natural change on the screen; there are no hidden states. To add an accent to any existing character, the user positions the insertion point (caret) after the character and types the accent."

Rendering non-spacing marks is the basic issue in the Urdu for atomization and combination. It is resolved in Unicode as (7:172-178):

"This discussion assumes the use of proportional fonts, where the widths of individual characters can vary. Various techniques can be used with monospaced fonts. In general, however, it is possible to get only a semblance of a correct rendering for most scripts in such fonts.

When rendering a sequence consisting of more than one non-spacing mark, the non-spacing marks should, by default, be stacked outward from the base character. That is, if two non-spacing marks appear over a base character, then the first nons-pacing mark should appear on top of the base character, and the second non-spacing mark should appear on top of the first. If two non-spacing marks appear under a base character, then the first nonspacing mark should appear

representing the combined form. In some cases the width of the base should change because of an applied accent, such as with "i". The ligature or contextual form mechanisms in the font can be used to change the width of the base in cases where this is required.

Correct multilingual comparison routines must already be able to compare a sequence of characters as one character, or one character as if it were a sequence. Such routines can also handle combining character sequences when supplied with the appropriate data. When searching strings, remember to check for additional non-spacing marks in the target string that may affect the interpretation of the last matching character.

Line breaking algorithms generally use state machines for determining word breaks. Such algorithms can be easily adapted to prevent separation of non-spacing marks from base characters.

A common implementation for the input of combining character sequences is the use of dead keys. These keys match the mechanics used by typewriters to generate such sequences through overtyping the base character after the non-spacing mark. In computer implementations, keyboards enter a special state when a dead key is pressed for the accent and emit a precomposed character only when one of a limited number of "legal" base characters is entered. It is straightforward to adapt such a system to emit combining character sequences or precomposed characters as needed.

Typists, especially in the Latin script, are trained on systems that work using dead keys. However, many scripts in the Unicode Standard (including the Latin script) may be imple mented according to the handwriting sequence, in which users type the base character first, followed by the accents or other

Modern systems indicate selection with some sort of rectangular highlighting. This approach places restrictions on the consistency of editing because some sequences of characters do not linearly progress from the start of the line. When characters stack, two mechanisms are used to visually indicate partial selection: linear and nonlinear boundaries.

Use of linear boundaries treats the entire width of the resultant glyph as belonging to the first character of the sequence, and the remaining characters in the backing-store representation as having no width and being visually afterward.

A relatively small number of implementation features are needed to support nonspacing marks. Different levels of implementation are also possible. A minimal system yields good results and is relatively simple to implement. Most of the features required by such a system are simply modifications of existing software.

As non-spacing marks are required for a number of writing systems, such as Arabic, Hebrew, and those of South Asia, many vendors already have systems capable of dealing with these characters and can use their experience to produce general-purpose software for handling these characters in the Unicode Standard.

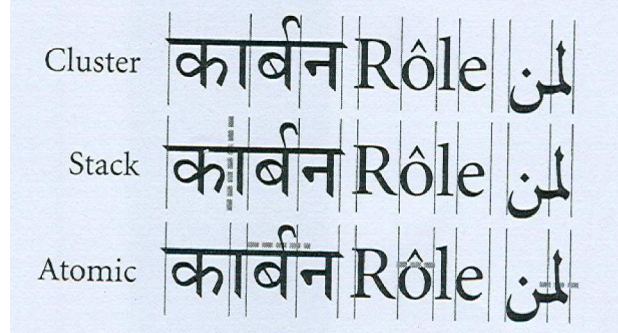
Composite character sequences can be rendered effectively by means of a fairly simple mechanism. In simple character rendering, a nonspacing combining mark has a zero advance width, and a composite character sequence will have the same width as the base character.

Wherever a sequence of base character plus one or more non-spacing marks occur, the glyphs for the non-spacing marks can be positioned relative to the base. The ligature mechanisms in the fonts can also substitute a glyph

user thinks of as a character and the key or key sequence used to input it.

Three types of boundaries are generally useful in editing and selecting within words: cluster boundaries, stacked boundaries and atomic character boundaries.

Cluster Boundaries. Arbitrarily defined cluster boundaries may occur in scrip\s such as Devanagari, for which selection may be defined as applying to syllables or parts of syllables. In such cases, combining character sequences such as ka + vowel sign a or conjunct clusters such as ka + halant + ta are selected as a single unit. (See Figure 5-4.)

Figure 5-4. Consistent Character Boundaries

Stacked boundaries are generally somewhat finer than cluster boundaries. Free-standing elements (such as vowel sign a in Devanagari) can be independently selected, but any elements that "stack" (including vertical ligatures such as Arabic lam + meem in Figure 5-4) can be selected only as a single unit. Stacked boundaries treat default grapheme clusters as single entities, much like composite characters.

The use of atomic character boundaries is closest to selection of individual Unicode characters. However, most I1)

a relatively small number of precomposed base character plus nonspacing marks have independent Unicode character values. Most existed in dominant standards.

Systems that cannot handle nonspacing marks can normalize to precomposed characters; this option can accommodate most modern Latin-based languages. Such systems can use fallback rendering techniques to at least visually indicate combinations that they cannot handle (see the "Fallback Rendering" subsection of Section 5.13, Rendering Nonspacing Marks)."

Editing and selection in another point of atomization to be dicussed by Unicode and resolved as (7:167-170).

"As far as a user is concerned, the underlying representation of text is not a material concern, but it is important that an editing interface present a uniform implementation of what the user thinks of as characters. (See "'Characters' and Grapheme Clusters" in Section 2.11, Combining Characters.) The user expects them to behave as units in terms of mouse selection, arrow key movement, backspacing, and so on. For example, when such behavior is implemented, and an accented letter is represented by a sequence of base character plus a nonspacing combining mark, using the right arrow key would logically skip from the start of the base character to the end of the last nons pacing character.

In some cases, editing a user-perceived "character" or visual cluster element by element may be the preferred way. For example, a system might have the backspace key delete by using the underlying code point, while the delete key could delete an entire cluster. Moreover, because of the way keyboards and input method editors are implemented, there often may not be a one-to-one relationship between what the

2-25 is for the Greek hooked upsilon symbol with an acute accent. It has a canonical decomposition to one sequence and a compatibility decomposition to a different sequence.

For more precise definitions of these terms, see Chapter 3, Conformance.

Alternative Spellings. The Unicode Standard contains explicit codes for the most frequently used accented characters. These characters can also be composed; in the case of accented letters, characters can be composed from a base character and nonspacing mark(s). .

The Unicode Standard provides decompositions for characters that can be composed using a base character plus one or more nonspacing marks. Implementations that are "liberal" in what they accept but "conservative" in what they issue will have the fewest compatibility problems.

The decomposition mappings are specific to a particular version of the Unicode Standard. Further decomposition mappings may be added to the standard for new characters encoded in the future; however, no existing decomposition mapping for a currently encoded character will ever be removed, nor will a decomposition mapping be added for a currently encoded character. This follows from the stability guarantees for normalization. See Appendix F, Unicode Encoding Stability Policies, for more information.

We may consider the Normalization as a serious issue, elaborated by Unicode as (7:160):

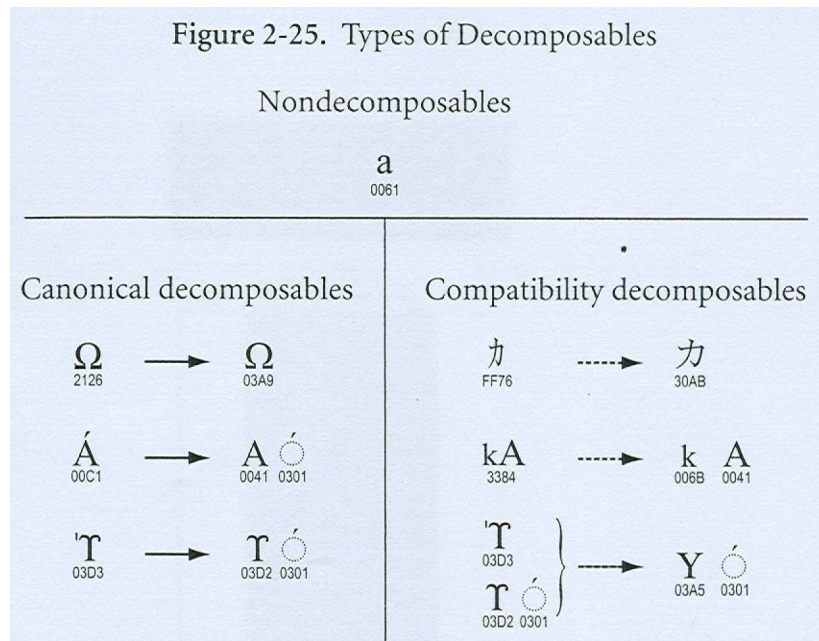
Systems may normalize Unicode-encoded text to one particular sequence, such as normalizing composite character sequences into precomposed characters, or vice versa.

Compared to the number of possible combinations, only

nondecomposables.

Figure 2-25 illustrates these three types. The solid arrows indicate canonical decompositions and the dotted arrows indicate compatibility decompositions. Compatibility decompositions that are redundant because they are identical to the canonical decompositions are not shown. The figure illustrates two important points:

Decompositions may be to single characters or to sequences of characters. Decompositions to a single character, also known as singleton decompositions, are seen for the ohm sign and the halfwidth katakana ka in Figure 2-25. Because of examples like these, decomposable characters in Unicode do not always consist of obvious, separate parts; one can know their status only by examining the data tables for the standard.

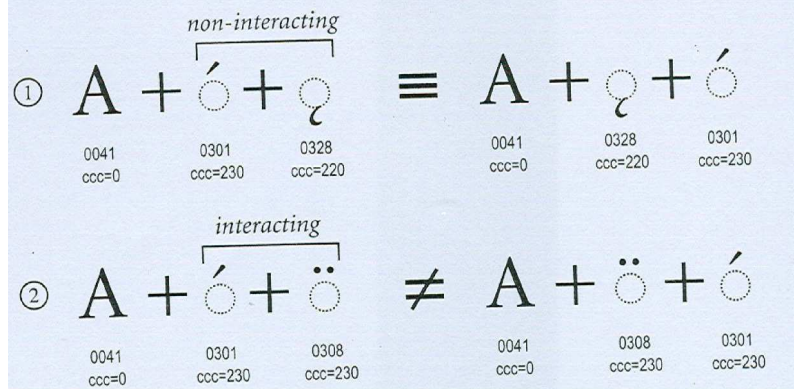
Figure 2-25. Types of Decomposables

A very small number of characters are both canonical and compatibility decomposable. The example shown in Figure

combining accent. The vast majority of nontrivial decompositions are indeed a mapping from a character code to a character sequence. However, in a small number of exceptional cases, there is a mapping from one character to another character, such as the mapping from ohm to capital omega. Finally, there are the "trivial" decompositions, which are simply a mapping of a character to itself. They are really an indication that a character cannot be decomposed, but are defined so that all characters formally have a decomposition. The definition of decomposable is written to encompass only the nontrivial types of decompositions; therefore these characters are considered nondecomposable.

In summary, three types of characters are distinguished based on their decomposition behavior:

- Canonical decomposable. A character that is not identical to its canonical decomposition.
- Compatibility decomposable. A character whose compatibility decomposition is not identical to its canonical decomposition.
- Nondecomposable. A character that is identical to both its canonical decomposition and its compatibility decomposition. In other words, the character has trivial decompositions (decompositions to itself). Loosely speaking, these characters are said to have "no decomposition;' even though, for completeness, the algorithm that defines decomposition maps such characters to themselves.

Because of the way decompositions are defined, a character cannot have a nontrivial canonical decomposition while having a trivial compatibility decomposition. Characters with a trivial compatibility decomposition are therefore always

Figure 2-24. Canonical Ordering

ascending order of the Canonical- Combining Class (ccc) values. The ccc values are shown below each character. The second row of Figure 2-24 shows an example where

combining marks interact typographically-the two sequences have different stacking order, and the order of combining marks is significant. Because the two combining marks have been given the same combining class, their ordering is retained under canonical.reordering. Thus the two sequences in the second row are not equivalent.

Precomposed characters are formally known as decomposables, because they have decompositions to one or more other characters. There are two types of decompositions:

. Canonical. The character and its decomposition should be treated as essentially equivalent.

. Compatibility. The decomposition may remove some information (typically formatting information) that is important to preserve in particular contexts.

A decomposition implies reducing a character to an equivalent sequence of constituent parts, such as mapping an accented character to a base character followed by a

differences in display and may be interpreted differently In some contexts.

If an application or user attempts to distinguish between' canonically equivalent sequences, as shown in the first example in Figure 2-23, there is no guarantee that other applications would recognize the same distinctions. To prevent the introduction of interoperability problems between applications, such distinctions must be avoided wherever possible. Making distinctions between compatibly equivalent sequences is less problematical. However, in restricted contexts, such as the use of identifiers, avoiding compatibly equivalent sequences reduces possible security issues. See Unicode. Technical Report #36, "Unicode Security Considerations."

Where a unique representation is required, a normalized form of Unicode text can be used to eliminate unwanted distinctions. The Unicode Standard defines four normalization forms: Normalization Form D (NFD), Normalization Form KD (NFKD), Normalization Form C (NFC), and Normalization Form KC (NFKC). Roughly speaking, NFD and NFKD decompose characters where possible, while NFC and NFKC compose characters where possible. For more information, see Unicode Standard Annex #15, "Unicode Normalization Forms;' and Section 5.6, Normalization.

A key part of normalization is to provide a unique canonical order for visually nondistinct sequences of combining characters. Figure 2-24 shows the effect of canonical ordering for multiple combining marks applied to the same base character.

In the first row of Figure 2-24, the two sequences are visually nondistinct and, therefore, equivalent. The sequence on the right has been put into canonical order by reordering in

scripts, such as Arabic, this situation occurs quite often when vowel marks are used. It arises because of the large number of ligatures in Arabic, where each element of a ligature is a consonant, which in turn can have a vowel mark attached to it. Ligatures can even occur with three or more characters merging; vowel marks may be attached to each part."

Figure 2-22. Ligated Multiple Base Characters

f + ◌̃ + i + ◌̣ → f̃ị

0066 0303 0069 0323

After our proposal a question of normalization arose which is already solved in Unicode as (7: 54-57):

In cases involving two or more sequences considered to be equivalent, the Unicode Standard does not prescribe one particular sequence as being the correct one; instead, each

sequence is merely equivalent to the others. Figure 2-23 illustrates the two major forms of equivalent sequences formally defined by the Unicode Standard. In the first example, the sequences are canonically equivalent. Both sequences should display and be interpreted the same way. The second and third examples illustrate different compatibility sequences. Compatible-equivalent sequences may have format

Figure 2-23. Equivalent Sequences

① B + Ä ≡ B + A + ◌̈ (0042 00C4 0042 0041 0308)

② LJ + A ≈ L + J + A (01C7 0041 004C 004A 0041)

③ 2 + ¼ ≈ 2 + 1 + ⁄ + 4 (0032 00BC 0032 0031 2044 0034)

demonstrates the script-dependent or writing -system dependent nature of rendering combining diacritical marks.

For additional examples of script-specific departure from

Table 2-6. Nondefault Stacking

Glyph	Sequence	
ἄ	GREEK SMALL LETTER ALPHA + COMBINING COMMA ABOVE (psili) + COMBINING ACUTE ACCENT (oxia)	This is correct
ά̓	GREEK SMALL LETTER ALPHA + COMBINING ACUTE ACCENT (oxia) + COMBINING COMMA ABOVE (psili)	This is incorrect

default stacking of sequences of combining marks, see the discussion of positioning of multiple points and marks in Section 8.1, Hebrew, or the discussion of nondefault placement of Arabic vowel marks accompanying Figure 8-5 in Section 8.2, Arabic.

The Unicode Standard specifies default stacking behavior to offer guidance about which character codes are to be used in which order to represent the text, so that texts containing multiple combining marks can be interchanged reliably. The Unicode Standard does not aim to regulate or restrict typographical tradition.

When the glyphs representing two base characters merge to form a ligature, the combining characters must be rendered correctly in relation to the ligated glyph (see Figure 2-22). Internally, the software must distinguish between the nonspacing marks that apply to positions relative to the first part of the ligature glyph and those that apply to the second part.

Ligated base characters with multiple combining marks do not commonly occur in most scripts. However, in some

When combining characters do not interact typographically, the relative ordering of contiguous combining marks cannot result in any visual distinction and thus is insignificant.

Another example of multiple combining characters abov;e the base character can be found in Thai, where a consonant letter can have above it one of the vowels U+OE34 through U+OE37 and, above that, one of four tone marks U+OE48 through U+OE4B. The order of character codes that produces this graphic display is base. consonant character + vowel character + tone mark character.

Many combining characters have specific typographical traditions that provide detailed rules for the expected rendering. These rules override the default stacking behavior. For example, certain combinations of combining marks are sometimes positioned horizontally rather than stacking or by ligature with an adjacent nonspacing mark (see Table 2-6). When positioned horizontally, the order of codes is reflected by positioning in the predominant direction of the script with which the codes are used. For example, in a left-to-right script, horizontal accents would be coded from left to right. In Table 2-6, the top example is correct and the bottom example is incorrect.

Such override behavior is associated with specific scripts or alphabets. For example, when used with the Greek script, the "breathing marks" U+0313 COMBINING COMMA ABOVE (psili) and U+0314 COMBINING REVERSED COMMA ABOVE (dasia) require that, when used together with a following acute or grave accent, they be rendered side-by-side rather than the accent marks being stacked above the breathing marks. The order of codes here is base character code + breathing mark code + accent mark code. This examrle

discussion summarizes the default treatment of multiple combining characters. (For the formal algorithm, see Section 3.11, Canonical Ordering Behavior.)

Table 2-5. Interaction of Combining Characters

Glyph	Equivalent Sequences
ã	LATIN SMALL LETTER A WITH TILDE LATIN SMALL LETTER A + COMBINING TILDE
ȧ	LATIN SMALL LETTER A WITH DOT ABOVE LATIN SMALL LETTER A + COMBINING DOT ABOVE
ạ̃	LATIN SMALL LETTER A WITH TILDE + COMBINING DOT BELOW LATIN SMALL LETTER A + COMBINING TILDE + COMBINING DOT BELOW LATIN SMALL LETTER A WITH DOT BELOW + COMBINING TILDE LATIN SMALL LETTER A + COMBINING DOT BELOW + COMBINING TILDE
ạ̇	LATIN SMALL LETTER A WITH DOT BELOW + COMBINING DOT ABOVE LATIN SMALL LETTER A + COMBINING DOT BELOW + COMBINING DOT ABOVE LATIN SMALL LETTER A WITH DOT ABOVE + COMBINING DOT BELOW LATIN SMALL LETTER A + COMBINING DOT ABOVE + COMBINING DOT BELOW
ấ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE LATIN SMALL LETTER A WITH CIRCUMFLEX + COMBINING ACUTE LATIN SMALL LETTER A + COMBINING CIRCUMFLEX + COMBINING ACUTE
á̂	LATIN SMALL LETTER A ACUTE + COMBINING CIRCUMFLEX LATIN SMALL LETTER A + COMBINING ACUTE + COMBINING CIRCUMFLEX

If the combining characters can interact typographically-for example, U+0304 COMBINING MACRON and U+0308 COMBINING DIAEREsIs-then the order of graphic display is determined by the order of coded characters (see Table 2-5). By default, the diacritics or other combining characters are positioned from the base character's glyph outward. Combining characters placed above a base character will be stacked vertically, starting with the first encountered in the logical store and continuing for as many marks above as are required by the character codes following the base character. For combining characters placed below a base character, the situation is reversed, with the combining characters starting from the base character and stacking downward.

character is an enclosing combining mark-in other words, when its GeneraLCategory value is Me-the resulting sequence has the appearance of a symbol. In Figure 2-19, enclosing the exclamation mark with U+20E4 COMBINING ENCLOSING UPWARD POINTING TRIANGLE produces a sequence that looks like U+26AO, WARNING SIGN.

Some Indic vowel signs are rendered to the left of a consonant letter or consonant ch~ster, even though their logical order in the Unicode encoding folloW6 the consonant letter. In the charts, these vowels are depicted to the left of dotted circles (see Figure 2-20). The coding of these vowels in pronunciation order and not in visual order is consistent with the ISCII standard.

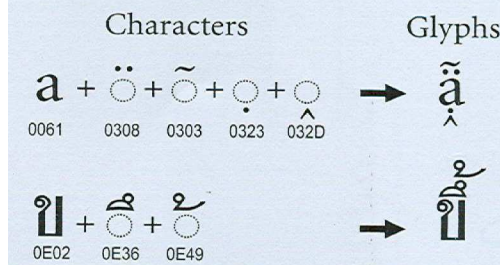
Figure 2-20. Reordered Indic Vowel Signs

फ + ि → फि

092B 093F

Our Ghost Characters Theory is descreteed in the Unicode as a function only i.e. Multiple Combining Characters (7:50-53):

"In some instances, more than one diacritical mark is applied to a single base character (see Figure 2-21). The Unicode Standard does not restrict the number of combining characters that may follow a base character. The following

Figure 2-21. Stacking Sequences

such, they are not generally referred to as diacritics, but may have script-specific terminology such as harakat (Arabic) or matra (Devanagari). See Section 7.9, Combining Marks.

In the Unicode Standard, all combining characters are to be used in sequence following the base characters to which they apply. The sequence of Unicode characters <U+OO61 "a" LATIN SMALL LETTER A, U+0308 ''{:';''COMBINING DIAERESIS, U+OO75 "u" LATIN SMALL LETTER U> unambiguously represents "au" and not "aft", as shown in Figure 2-18.

Figure 2-18. Sequence of Base Characters and Diacritics

a + ◌̈ + u → äu (*not* aü)
0061 0308 0075

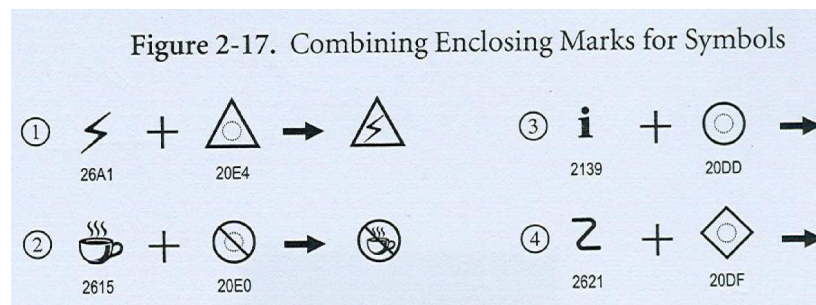
The ordering convention used by the Unicode Standard-placing combining marks after the base character to which they apply-is consistent with the logical order of combining characters in Semitic and Indic scripts, the great majority of which (logically or phonetically) follow the base characters with respect to which they are positioned. This convention also conforms to the way modern font technology handles the rendering of nonspacing graphical forms (glyphs) so that mapping from character memory representation order to font rendering order is simplified. It is different from the convention used in the bibliographic standard ISO 5426.

A sequence of a base character plus one or more combining characters generally has the same properties as the base character. For example, "N' followed by "A" has the same properties as "A': For this reason, most Unicode algorithms ensure that such sequences behave the same way as the corresponding base character. However, when the combining

other atoms, e.g. diatrictics, symbols and other characters to form letter. It allows double, triple code value to a single letter like آ has Unicode vlaue of 0662 and if it is considered a combination of Alif (ا) (0627) and Mudd (~)(0653) then we can add both the Unicode values to آ. We can process آ both the ways Single letter and key and atomised characters keys. This is shown in unicode as (7:48-50):

Figure 2-17 shows examples of combining enclosing marks for symbols. The combination of an enclosing mark with a base character has the appearance of a symbol. As discussed in "Properties" later in this section, it is best to limit the use of combining enclosing marks to characters that encode symbols. A few symbol characters are intended primarily for use with enclosing combining marks. For example, U+2139 INFORMATION SOURCE is a symbol intended for use with U+20DD COMBINING ENCLOSING CIRCLE or U+20E2 COMBINING ENCLOSING SCREEN. U+2621 CAUTION SIGN is a winding road symbol that can be used in combination with U+20E4 COMBINING ENCLOSING UPWARD POINTING TRIANGLE or U+20DF COMBINING ENCLOSING DIAMOND.

Some scripts, such as Hebrew, Arabic, and the scripts of India and Southeast Asia, have both spacing and nonspacing combining characters specific to those scripts. Many of these combining characters encode vowel letters. As

Figure 2-17. Combining Enclosing Marks for Symbols

characters.

According to the proposal, you were requesting 22 additions. I have taken the names and tried to align them with the notion of being spacing characters. The names are also updated to the usual style for such characters, beginning with "Arabic" for the script, and then annotated where appropriate for the particular language." (See appendix).

0880 ARABIC SINGLE NUQTA ABOVE
0881 ARABIC SINGLE NUQTA BELOW
0882 ARABIC DOUBLE NUQT A ABOVE
0883 ARABIC DOUBLE NUQTA BELOW
0884 ARABIC TRIPLE NUQTA ABOVE
0885 ARABIC TRIPLE NUQTA BELOW
0886 ARABIC TRIPLE INVERTED NUQTA ABOVE
0887 ARABIC TRIPLE INVERTED NUQTA BELOW
0888 ARABIC QUADRUPLE NUQTA ABOVE * Sindhi
0889 ARABIC QUADRUPLE NUQTA BELOW * Sindhi
088A ARABIC DOUBLE DANDA ABOVE * Sindhi
088B ARABIC DOUBLE DANDA BELOW * Sindhi
088C ARABIC DOUBLE NUQT A VERTICAL ABOVE * Sindhi
088D ARABIC DOUBLE NUQTA VERTICAL BELOW * Sindhi
088E ARABIC SINGLE KASHIDA ABOVE * Urdu
088F ARABIC SINGLE KASHIDA BELOW * Urdu
0890 ARABIC DOUBLE KASHIDA ABOVE * Urdu
0891 ARABIC DOUBLE KASHIDA BELOW * Urdu
0892 ARABIC SINGLE CIRCLE ABOVE *Pashto
0893 ARABIC SINGLE CIRCLE BELOW *Pashto
0894 ARABIC TOTA ABOVE * Urdu
0895 ARABIC TOTA BELOW * Urdu

Unicode already allowed Ghost Characters to combine

candidates for such a proposal.

In the meantime, as regards the nuqta letters, the main points of controversy within UTC revolved around the confusion that would be created if a set of formally non-spacing characters were added for the composition of new letters, as Mr Kew outlined. And this would also create destabilization of the standard. However, in discussion, several members of the UTC seemed to agree that there is a useful path forward with the addition of formally spacing dot patterns that can be used for pedagogical and other purposes, as exemplified in your proposal.

On our end, we need to look carefully at the correct semantics and character property assignments for such a set of dot patterns, and we need to agree on the exact repertoire of letters that will be required. And we need to be able to accurately document the letters and their method of use so that we avoid confusion in the community of implementors.

Your original proposal L2/06-039 listed 22 characters to be encoded within the 0600 block. Several people have suggested that the open slots on that block should be reserved for further symbols and full letters, and that the dot patterns should be proposed for encoding in another block. There is an open area of right-to-left codepoints tentatively reserved for Arabic additions in the range from 0880 through 08BF (and possibly more). This range is probably the best place to put a set of spacing nukta

> such characters were added in Unicode 4.1, and at this last UTC meeting seven more characters were approved for addition in Unicode 5.1. We would welcome efforts by you or others to conduct in-depth investigations to determine as soon as possible what additional Arabic-script characters are in use, which need to be added to the Unicode Standard. That won't eliminate an ongoing need to add new characters, but hopefully it would lead to implementations for a wide range of language communities in the not-too-distant future."

After a long discussion and suggestions made by me Mr. Rick McGowan of Unicode wrote to this worthy proposal that (10:3):

> "The proposal L2/06/039 was controversial when presented to the UTC in 2006, and it has taken some time to secure the assistance of someone here to work with us on a new proposal that we hope will satisfy the need for encoded dot patterns, and be acceptable to UTC.
>
> Jonathan Kew wrote to you in July 2006 after the initial UTC discussion of your proposal L2/06-039. In that letter he covered two topics. The first topic regards the reasons why it is impossible to change the model for Arabic encoding at this stage to include non-spacing equivalents for the existing encoded characters. The second topic was an invitation to submit further characters required for the minority languages of Pakistan. I hope that we can work together in the coming months to gather

Unfortunately, the UTC does not have much flexibility on this matter due to externally imposed demands for stability. In your response to Jonathan Kew, you indicated that the intent is not to change the encoding model for Arabic or to introduce an additional encoding model parallel to the existing one. Yet that is exactly what addition of these combining nuqta characters would do, and that change would constitute an untenable destabilization in the Unicode Standard. It was for precisely that reason that UTC was forced to reject the proposal in 2003, even though the committee as a whole agreed that a decomposed representation for Arabic script would have been preferable had it been done from the outset before stability became a limiting factor.

Please note that the limitation on nuqtas used to create additional "letters" does not apply to Quranic marks that are added as independent annotations or ornamentations to the Quranic text. Combining marks of this type can be added in order to provide complete encoded representation of Quranic text. Indeed, four such combining marks were approved for encoding at this last UTC meeting.

We recognize the real concern this creates for language communities that require Arabic letter-nuqta combinations that cannot presently be represented in Unicode. For this reason, we particularly appreciate proposals such as those that have been made since 2003 to add additional characters for such languages. Thirty

> the FAQ mentioned in the Open Action Item 108-A2 designated to Lisa Moore."

Mr. Mark Davis of the Unicode created FAQ and replied on March 30, 2007 that (10:2):

> "On behalf of the Unicode Technical Committee (UTC), I would like to thank you for your contributions regarding the proposed addition of combining nuqta characters (L2/06039, L2/06-322). These contributions were considered at the recent UTC meeting held November 7 to 9 in San Jose.
>
> As Jonathan Kew indicated in his response to your initial contribution, a decision was taken by UTC in 2003 specifically not to encode combining nuqta characters. At the recent UTC meeting, the committee reaffirmed that earlier decision.
>
> We recognize that this is, in several respects, less than ideal. The committee understands the potential benefits that separating the skeletal consonant letters from the various nuqtas could have. In particular, we are aware that usage of the Arabic script is continuing to evolve as more and more languages adopt this script as the basis of practical orthographies, often requiring innovation of new combinations of skeletal letters and nuqtas. This was a key aspect of the proposal to encode nuqtas made in 2003. There is general consensus among UTC members that a decomposed representation model for Arabic script would have been preferable all along.

canvas expanded, ASCII was short enough to even support all the European languages and the reason was simple: ASCII was not considering the Latin Script rather just English (7 bits) and few European languages (8 bits). But later on, when the whole stage for the languages was considered, decomposed marks were to be added. Same is the case with Arabic. The motive behind the initial digitization of Arabic was to capture the Middle East Markets. Thus vendors altogether ignored the rest of the languages of Arabic script having a narrow view about the script developing proprietary standards covering only Arabic and at most Persian. Arabic coverage of Unicode is simply reflection and continuation of that approach which is definitely inadequate for the future needs of the script

Lastly, regarding your mention about the font development using OpenType, we've already developed a Nastaleeq font (named Pak Nastaleeq) which stands as our National Standard supporting all the regional languages whether digitized or not and we've deployed the same techniques you described of decomposing at the glyph level. But to support this idea we need the addition of Nuqta marks into the Arabic Code block. This font is likely to be adopted by the Urdu versions of the Microsoft products replacing the traditional Naskh style screen layouts into Nastaleeq as we are in close negotiations with Microsoft as their Localization Program Partner regarding this.

The facts described herein can be rephrased into

costs involved but the associated benefits must be considered as well. Arabic script has three major languages: Arabic itself, Urdu and Persian. Currently, the amount of electronic text in these languages is not that much enormous as compared to the languages written in Latin script like English, French etc. So a radical change at this stage is inevitable otherwise delay will further make it impossible to impose any restructuring on the Arabic Block thus yielding a wrong direction.

Definitely, the idea of combining marks generating new characters is not new to the Unicode.

Many existing scripts covered by the Unicode enjoy the support for combining diacritics marks that sit of the base characters in order to form new characters. Latin is just an example of that (See Section 7.7 from The Unicode Standard 4.0). Indic group of languages have also combining marks associated with them. Devanagri has such a code point (U+093C DEV ANAGRI SIGN NUKTA) "for extending the alphabet to new characters" as per the documentation in the standard.

Current state of the Arabic Code block is somewhat analogous to the ASCII. ASCII was developed by IBM as their proprietary standard on their machines using 7 bits but it got wide acceptance getting an expansion of 8 bit including many characters from European languages. But later on, as the computing

> script. Now it is complimentary to add support for these Nuqta Characters to the Arabic code block to realize the real benefit of the set of Ghost Characters."

Moreover, Nuqtas are also present in Quran as separate characters like 2, 3 and 4 Nuqtas above used separately. In these circumstances, need for these Nuqta marks as separate characters is of immense importance Mr. Jonathan argued (10:1).

> Regarding the current encoding model followed by Unicode about Arabic script, we do not intend to change the system or introduce a parallel or duplicate encoding system in the Arabic Block. Merely, what we want is the addition of these Nuqta characters along with the proposed properties. If it introduces a parallel system then it's an additional benefit yielding self sufficiency of the Arabic script.
>
> However, a point that we would like to make is that enormous numbers of languages are adopting Arabic Script as the script of their choice and are defining their own unique ways to represent sounds not available in the Arabic phonetic system. Usually, registering a character to Unicode is a lengthy process spanning over several years and this waiting queue merely results in take-off delay of these languages into the digital age. But the presence of these proposed characters will provide somewhat "Grand Unification Theory" like effect to the Arabic script. Regarding your mention about costs, definitely every transition has associated

alternate for four dots.

Finally, if there are additional diacritic marks that are used as diacritics, similarly to the Arabic vowel marks, Qur'anic annotation marks, etc., these can of course be considered for addition to Unicode. But if they are used to create new Arabic-script consonants, along the pattern of the nuqtas and small tah, then such letters should be proposed as individual characters."

I considered the letter of Jonathan Kew in letters and spirit but the proposal remained. Then I replied to Mr. Jonathan that (10:1):

> "I received your comments regarding our proposal for addition of Preliminary Proposal to add Nuqta Characters to Arabic Block (L2/06-039 now moved to L2-06-240). Reposing our stance regarding the Nuqtas, These characters have individualistic script existence and are often needed in the generation of electronic texts like pedagogical material. You will find scanned images of two primers published by the Pakistani provincial governments in wide circulation which use Nuqtas as separate characters. In situations like these, we have to resort to the old fashioned way that is the calligrapher leaving our DTP systems of no use. And definitely this introduces much more complications in the electronic storage, retrieval and transmission of these texts. Previously, Unicode added many entries from our ASCII Code Plate notification including the notion of Ghost Characters thus completing the set of atoms of Ghost Characters for the Arabic

as base characters for the construction of modern dotted letters.

It is important to remember that the use of a "precomposed" encoding model for Arabic letters in Unicode does not prevent font developers using decomposed forms and dynamic composition technology to actually render the text. In complex styles such as Nastaliq, a typical OpenType font implementation may begin with a glyph decomposition step, mapping each letter to a base form and one or more combining mark glyphs; then contextual replacement and positioning rules can be applied to these. But this happens entirely within the font, using font-specific glyph codes, and is not related to the Unicode encoding.

To conclude, then, the combining nuqta characters are not regarded as valid candidates for addition to the Standard. That would be a very reasonable model for encoding the Arabic script, but it is not the model used in Unicode.

In the light of this, I would invite you to submit proposals for any specific characters required that are not currently available in Unicode, whether these are for regional languages or archaic documents. In particular, I think the reh with four dots below shown in Figure 2 of your document is not yet encoded, and there may be others as well.

In the case of the double lines above or below letters, used with seen and hah base shapes, I think that these may simply represent alternate forms, based on hand-writing, of a four-dot combination. In Shina, at least, I am familiar with the use of both seen with four dots above and seen with two lines above to represent a retroflexed Ṣ, and I suspect that the same may be true for ch written as hah with four dots below. It would be debatable, therefore, whether the two-line combination should be encoded separately, or merely regarded as a glyph

normalization forms. These proposals were considered by the. UTC in June 2003. After considerable discussion, the final decision was that even though such an encoding model would have real benefits, the costs of making such a change at this stage would be too great. I think this related both to technical (implementation) requirements and to the confusion that introducing a parallel encoding model could cause. Ultimately, motion 95-M1 In principle the UTC is infavor of encoding additional diacritic combining characters for productive extension of the Arabic script failed to pass, and I do not think the committee is likely to reconsider this question now.

This is, then, the clear policy in Unicode: Arabic letters are encoded as indivisible units, including their nuqta combinations, and are not composed dynamically. As a result, new letters continue to be added to the Standard, as documentation is presented.

ٮ

In passing, note that there are some cases where an Arabic-script letter is visually similar to a base letter plus an existing diacritic such as U + 0654 hamza above or U + 0615 small high tah. In cases like U + 0623 أ or U + 0626I ئ.>, where the meaning is indeed a letter with hamza added, these characters have canonical decompoݙtions. But in the case of U + 0681 ځ U + 076C ٮ or U + 0759 where the hamza or tah shape is used as part of a new letter, not as a separate letter or mark in its own right, there is no decomposition; these characters are encoded as distinct letters.

Your document also questioned the purpose of encoding the dotless forms U + 066F ب and U + 066F 0, when there are no combining nuqta marks that can be added to form complete letters. These were encoded to complete the set of dotless skeletal letters so that archaic undotted Arabic text can be represented in Unicode; they are not expected to function

Therefore, if a set of combining nuqta characters were added, there would be two ways to encode most Arabic-script letters: the old individual characters, or the new base + nuqta sequences. This is not good for users, as it would lead to confusion and ambiguity, and to problems such as spoofing.

Where Unicode does, in fact, have two ways to encode a given piece of text (accented Latin characters being a primary example), the concepts of canonical equivalence and normalization are there to solve most of the problems this could cause. Thus, Unicode specifies that U+0101 ā has a canonical decomposition to <U+0061 a, U+0304 , ◌̄ >. This means that all systems can treat these representations as equivalent, and convert between them as needed.

It is natural to suggest, then, that the existing Arabic letters should be given canonical decompositions to base + nuqta sequences, just as the accented Latin letters have. Thus, U +062A ت would have a decomposition to <U +066E 'ٮ', U +XXXX ◌̈ >, etc. However, stability considerations do not allow existing characters to be given decompositions where they do not already have them. Any change that would disrupt normalization stability could have bad effects on other standards and protocols that depend on Unicode, and could lead to major interoperability problems between systems. Therefore, the "obvious" solution is not available to us.

Several years ago, in fact, I co-authored a proposal, with input from others on the Committee, for a set of characters similar to the nuqta characters you have requested. At that time, a proposal was also presented for extending the normalization process. So that it would be possible to define a canonical equivalence between the existing Arabic letters and their base + nuqta representations, without destabilizing the

This document presents a concept that has been considered in the past: that Arabicscript letters could be encoded as combinations of "skeletal" letterforms such as 'ٮ' andح.

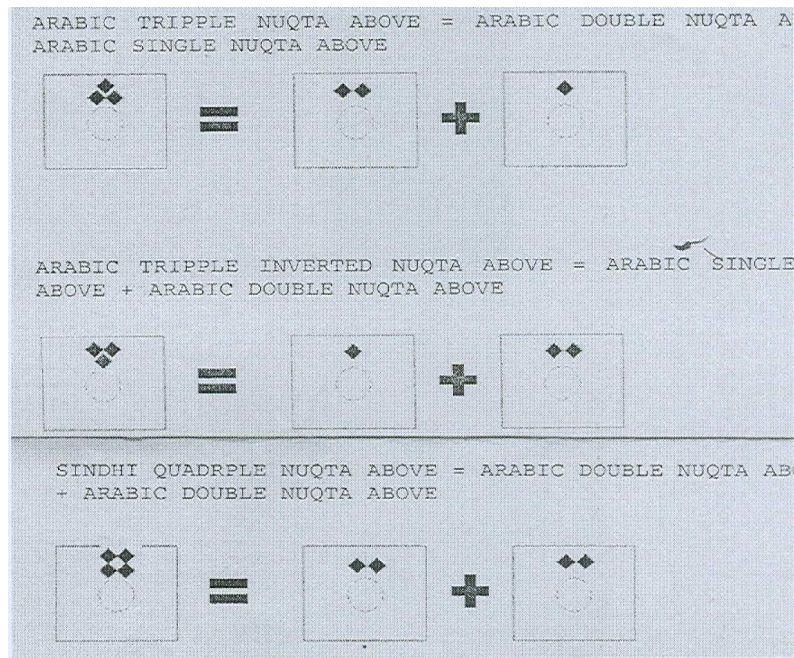
With dots (or sometimes other marks) positioned above or below, such as ◌ٔ , ◌ٖ or ◌ؕ. It is understood that this would be consistent with the inherent structure and historical development of the Arabic script, and would be helpful where regional languages are adopting new ways of writing using combinations of basic letter plus dots that have not previously been used. This approach can be referred to as a "generative" encoding of Arabic.

However, the established encoding practice in Unicode/ISO 10646 is that Arabic letters are encoded individually, as non-decomposable units. This practice predates the Unicode standard, which simply followed previous standards (particularly ISO 8859-6) in creating an initial repertoire of Arabic characters. The original repertoire (for Arabic, Persian, Urdu, etc.) has been supplemented with many new characters, as information becomes available.

Many experts now agree that it would have been preferable to use a generative model for Arabic; but at this time, it is not possible to re-design the encoding model used in the Unicode standard. There are a number of reasons why this is so.

Both practical considerations (the amount of existing data and software, all based on the current model) and published stability policies make it impossible to replace the existing Arabic letters with a base + nuqta encoding. Existing characters cannot under any circumstances be removed from the Standard.

Property	Value
Block(bc)	ARABIC
Bidi_Class(bc)	Nonspacing_Mark(NSM)
Canonical_Combining_Class(ccc)	Nukta(NK)
Joining_Group(jg)	NO_JOINING_GROUP
Joining_Type(jt)	Transparent(T)
Script(sc)	Arabic(Arab)

Further more, Nuqta characters having further canontical decompositions are list along with there canonical decompositions. We only present the above Nuqta can case for below Nuqta is Indentical to be above one.

On July 19, 2006 Mr. Johthan Kew from UNICODE wrote and arranged as (6&10):

" I am writing at the request of the Unicode Technical Committee with comments in response to your document L2/06-039 Preliminary Proposal to add Nuqta Characters to Arabic Block.

and is of historic nature. Other important reference is the dictionary published by Anjuman-e- Taraqqi-e-Urdu (Pakistan) in 1985, which was edited by the late Molvi Abdul Haq (also known as Father of Urdu, Baba-e-Urdu) (2).

Farhang-e- Talaffuz is another important document published by National Language Authority (Government of Pakistan) (3).

Sheena Quaida is an educational text published by Sarhad Provincial Government (Government of Pakistan) with support from Himalay Jungle Project financed by European Community and British High Commission Islamabad (4) and Balti Quaida published by National Language Authority (5).

Moreover, Government of Pakistan is developing a standard Nastaleeq font usable for all the regional language. This font is based on the idea of Character Composition and hence to be practical, needs support of Nuqta in the Unicode Standard.

Detailed and formal properties of these characters will be provided after the initial response from the UTC.

The Unicode technical committee was presented with a proposal (document number L2/06-039) to add Nuqta marks to the Unicode Arabic Block on 6th February 2006. That document contained the characters to be added and their rational for addition to the Arabic Block. .

The intent of this document was to provide further detailed character properties and elaborate their status, usage and advantage of their presence in the Arabic Block.

Almost all the proposed characters posses same character properties so for the sake of simplicity, all are listed in the table below:

20. Pashto Single Circle Below
21. Urdu Letter Tota Above
22. Urdu Letter Tota Below

This way, not only we are able to support all the characters of all languages in Arabic script, but even if some appear later on, we are already future compatible.

Even today, from pedagogical point of view, the different characters are memorized by teaching their basic shape and the number of Nuqtas with their orientation.

The inclusion of these combining characters to the Arabic Block will simplify the representation of texts in Arabic scripts. Not only existing characters can be defined as a character sequences but new characters also can be formed which are even not yet encoded into the standard.

The Arabic script is mostly being used by the societies which are evolving and hence their scripts are open ended. New characters are being defined on a rapid pace for many regional languages but encoding all of them to the Unicode Standard is inconvenient. Such an approach will merely stuff the Arabic Code Block very rapidly and will require extensions to the block on almost yearly basis.

The presence of U+066E ARABIC LETTER DOT LESS BEH, U+066F ARABIC LETTER DOT LESS QAF and U+06A 1 ARABIC LETTER DOT LESS FEH is useless in the absence of these combining characters.

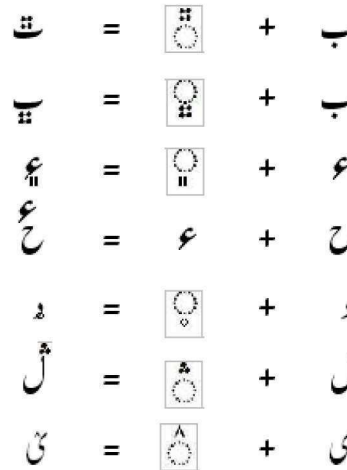
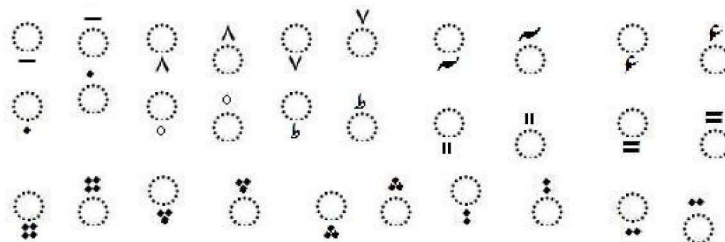
It is requested to provide code points for the above mentioned combining characters to simplify the process of Arabic Script processing.

Below some references are given. The first reference is about 300 years old (Muqalaat-e-Hafiz Mehmood Sheerani)

annex.

As an initial step some atoms of Urdu letters were placed in the ASCII code plate version 2 of NLA prepared in 1999 and exhibited in ITCN Asia 2001, Karachi., UNICODE Consortium took this initiative inclusive of Bey, Fey, Qaf without dots to place them in its version 3.1 and in published version 4.00 in Arabic Block but it ignored dots and other diacritics. A proposal was sent to the UNICODE consortium on 29-1-2006 to add dots or nuqta characters. Some 22 new combining Characters were to be added in the Arabic Block of Unicode. The addition of these combining marks will not only simplify the Arabic Block of Unicode Standard, but will make it possible to typeset all origional languages written in Arabic script. It was said a complimentary addition. These were:

1. Arabic Single Nuqta Above
2. Arabic Single Nuqta Below
3. Arabic Double Nuqta Above.
4. Arabic Double Nuqta Below
5.Arabic Tripple Nuqta Above
6. Arabic Tripple Nuqta Below
7. Arabic Tripple Inverted Nuqta Above
8. Arabic Tripple Inverted Nuqta Below
9. Sindhi Quadrple Nuqta Above
10. Sindhi Quadrple Nuqta Below
11. Sindhi Double Danda Above
12. Sindhi Double Danda Below
13. Sindhi Double Nuqta Vertical Above
14. Sindhi Double Nuqta Vertical Below
15. Urdu Single Kashida Above
16. Urdu Single Kashida Below
17. Urdu Double Kashida Above
18. Urdu Double Kashida Below
19. Pashto Single Circle Above

This is the basic theme of Atomization or combination. All other languages like Sindhi, Pashto, Balti, Balochi, Shina etc. use the same technique to develop their letters e.g.:-

It means that there are atleast 30 dots and diacratics as atoms that combine to the basic ghost characters to form a letter in Arabic script to be used by any language. These are in upper and lower placements as follows:

A total of 52 (22 + 30) atoms or characters form at least 660 different letters. These 660 can add one more ghost atom from dots and diacritics to form 19800 formal letters. The possibility of combining atoms to form letters goes up to a number of 20460 (600+ 19800 = 20460). It is a futuristic approach. If some one needs to see all these letters, it will require 80 pages of UNICODE having 256 letters on one page. Some of these possible 20460 letters may be seen in the

Urdu sounds (Phonemes) were to be derived adding four nuqtas in the ghost characters like: ٿ ڐ ڙ

These again converted the diacratics and dots having two dots and one line like: ٹ ڈ ڑ

After a long calligraphic practice the final shape of these letters in Urdu bacame in shape like this: ٹڈڑ

The Hindi phonemes or diphthongs were depicted with "ہ" like بہ، پہ، تہ، ٹہ جہ، چہ، دہ، ڈہ، رہ، ڑہ، کہ، گہ

The final shape is now using "ھ" character shape like: بھ، پھ، تھ، ٹھ، جھ، چھ، دھ، ڈھ، رھ، ڑھ، کھ، گھ.

Another "Yeh" "ی" was written in Urdu with half "ی" and then was converted into "ے".

It means that in Persian and Urdu languages three basic ghost characters were added, e.g. گ، ھ، ے

A total number of 22 ghost characters are now in use in all the languages using arabic script. These are as follows:

ا ٮ ح د ر س ص ط ع ڡ ٯ ک گ ل م ں و ہ ء ی ھ ے

Any letter of any language written in Arabic script has only 22 ghost characters in its basic system. The dots (Nuqtas), dandas, Toey and other diacritics or atoms are added for a combination or normalization to form any letter like:

ب = ◌ + ٮ

ٹ = ◌ + ٮ

ٹ = ◌ + ◌ + ٮ

ٹ = ◌ + ٮ

15. م=م

16. ن=ں

17. ہ=ہ

18. و ؤ وَ=و

19. ی ئ یَ=ی

The philosophy behind to dot the ghost characters, as coined by Ibne Maqla, was this that the first letter will have one dot, then two or three dots; first letter will have lower dots and then upper, may be introduced like:

ب ت ث

ج ح خ

د ذ

ر ز

س ش

ص ض

ط ظ

ع غ

ف ق

ں ن

Persia adopted Arabic script after the expansion of Islamic teachings and added some letters having three dots or one more line e.g.: پ(ب) چ(ج) ژ (ز) گ(ک)

When this script came into use in India, some Hindi

letters), so why Hajaj Bin Yousuf introduced dots to differenciate different letters emerging from the same roots of ideographs: one, two or three dots above or below.

The ghost (Khali) characters or common shapes of letters, we may call them "Kashties", were 19 in number. Holy Quran was inscribed in these Khali characters. The letters of the Holy Prophet (PBUH) were also written in these Kashties. The Arabs had no difficulty to read unwritten atoms (dots and diacritics) to utter the sound of any letter. These 19 characters were as follows:

1. ء
2. ا،أ،آ=ا
3. ب ت ث = ٮ
4. ج ح خ = ح
5. د ذ = د
6. ر ز = ر
7. س ش = س
8. ص ض = ص
9. ط ظ = ط
10. ع غ = ع
11. ف = ڡ
12. ق = ٯ
13. ک = ک
14. ل = ل

10

GHOST CHARACTERS: Atomization- Combination Theory

Phoenicians gave us the letters to write or record our messages in ideographic form. Before this innovation there were atleast two styles of writing systems: (1) Cuniforms of Summarians: using arrows for writing and (2) Hieroglyphic of Egyptians: using pictures to depict messages. The Phoenicians develped some basic signs: a conversion of pictorial elements into outlines to depict phonemes. These outlines were converted into two types of scripts (1) Arabic and (2) Roman, emerging from the same roots. CJK ideographs are another different story.

In Arabic script there are 29 letters derived from the Phonecian alphabet e.g.:-

ء ا ب ج د ھ و ز ح ط ی ک ل م ن س ع ف ص ق ر ش ت ث خ ذ ض ظ غ

Initially the Arabs were not using dots and diacritics to write the letters with these ideogrphs. The dotless arabic script was depicting basic letters like the following:

ء ا ٮ ح د ھ و ر (ح) ط ی ک ل م ں

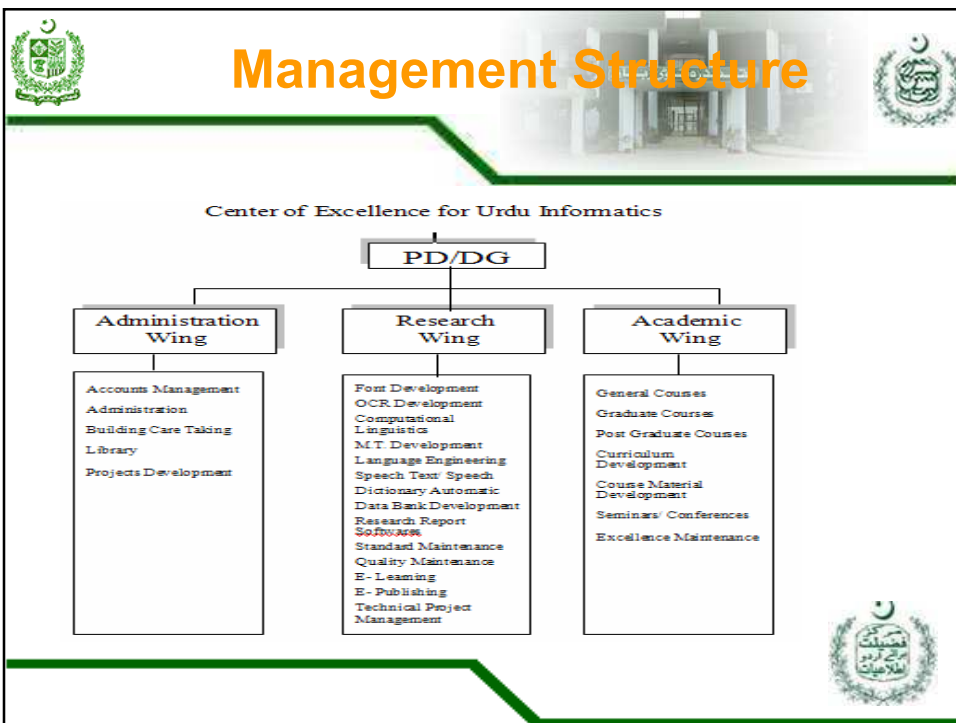
س ع ڡ ص ٯ (ر س ٮ ٮ ح د ص ط ع)

It means that they were using only 19 ghost characters and reading these letters by their cultural habit of usage and connotations. It were the non-arabs who could not become at home with these ghost characters or "Khali Kashtian" (Dotless

Proposal for an Independent Urdu Informatics Institute

National Institute for Urdu Informatics
(Center of Excellence for Urdu Informatics)

PC-I
Budgetary Details/ Estimates

S. No.	Details	Amount
1.	Civil Work	
	(a) Main Building 5 Storey 30K sq ft.	48,000,000
	(b) Hostel, Guest House, Base+ 2 storey, 20K sq/ft	12,000,000
	(c) Showroom- 2 storey	8,000,000
2.	Decor	16,500,000
3.	Chiller/ Air Conditioning	6,000,000
4.	Land Scaping	12,000,000
5.	Furnishing	4,000,000
6.	Application (Hard & Soft)	27,000,000
7.	Equipments, Connectivity etc.	4,000,000
8.	Training of Staff, inland, foreign	3,500,000
9.	Seminar/ Conferences	5,000,000
10.	Staffing	70,000,000
11.	Transport	12,000,000
12.	Utilities , POL etc., Installation, bills etc.	10,000,000
13.	Revolving Fund for future	20,000,000
14.	Library	2,500,000
14.	Annual Reoccurring	20,000,000
15.	Misc.	
	TOTAL	280,500,000

Relevance of CEUI

Organizations in Pakistan

- **Center of Excellence for Urdu Informatics (CEUI),** Pakistan
  Conducting research and development for Urdu standardization
  The Center of Excellence for Urdu Informatics (CEUI) is the Urdu IT wing of National Language Authority. CEUI conducts research and development activities in all matters relating to Urdu standardization for computers and localization. The main interests of the center are to:

- formulate and present recommendations to the Government for formal adoption of Urdu
- develop linkages for cooperation and collaboration with various academic, research and educational institutions
- consider ways and means for the promotion of Urdu as the National Language of Pakistan
- facilitate introduction of Urdu as the Official language throughout the country by developing dictionaries and other reading materials for in-service training of the personnel working in Government/ Semi-Government Offices, Courts and other Institutions

30 August 06: Institute for Urdu Informatics soon

ISLAMABAD: Institute for Urdu Informatics is an important need of the era and it will be soon established. It is the need of the time that an institute of Urdu Informatics in public sector should be enhance and boost the localization activities already being rendered in the world of business. Cabinet Secretary Ejaz Raheem expressed these view at display of “Pak Nastaleeq” font developed by Center of Excellence for Urdu Informatics, National Language Authority Islamabad. He was addressing ceremony before the initial or beta release of Unicode character based highly efficient “Pak Nastaleeq” font says a press release issued here in Monday. He said that the Prime Minster has decided to move the Urdu quickly into the cyber world. After the successful development of Urdu Office and Urdu Windows in collaboration with the Microsoft, he admired the development of Pak Nastaleeq font which is 90 times faster than other Nastaleeq fonts.

A New Field Established
Urdu Informatics
From Wikipedia, the free encyclopedia

- **Urdu Informatics** (Urdu: اردو اطلاعیات) relates to the cutting-edge research and efforts in bringing the utilities and usage of Urdu language to the modern information and communication technologies in education and businesses.
- Urdu Language Authority has been at the forefront in introducing Urdu Informatics as a tool for wider standardization of the language. Apart from development of Urdu keyboard, one of the key steps in this respect has been the establishment of a Centre of Excellence for Urdu Informatics in Islamabad, Pakistan.
- Works in Urdu Informatics also relate to the use of Urdu for internet applications, input controls, dynamic association of keyboard buttons to Urdu alphabets and use of proper character coding schemes.[1]

Notes

- ^ M.Z. Aziz, M.F. Aziz and K. Rashid (2004), *Converting Standard HTML Input Controls to Urdu*, Information Technology Journal, 3(1): pp. 69-73. Asian Network for Scientific Information. ISSN 1682-6027

- Vision
- Master, M.Phil and Ph.D Course are already in progress in the Department. We intend to start Diploma and Certificate Course for the people other than from Pakistan, and also working on the Degree Courses of Urdu informatics in near future.

سفارشات نفاذ اردو بطور دفتری زبان

Adoption of

Urdu as Official Language

Cabinet Committee Meeting 01-02-2007

(Decisions)

1۔ ڈاکٹر محمد وصی ظفر۔ وزیر قانون و انصاف

2۔ جناب محمد علی درانی- وزیر اطلاعات و نشریات

3۔ محترمہ زبیدہ جلال- وزیر سماجی بہبود و خصوصی تعلیم

4 ۔ ڈاکٹر عطاء الرحمن۔ چیئرمین HEC

فیصلے کابینہ کمیٹی یکم فروری 2007ء

1. مرکزفضیلت برائے اردو اطلاعیات (وفاقی اردو یونیورسٹی میں الحاق کے حوالے سے) HEC سے رقوم کے حصول کے لیے ترقیاتی منصوبہ تیار کیا جائے۔

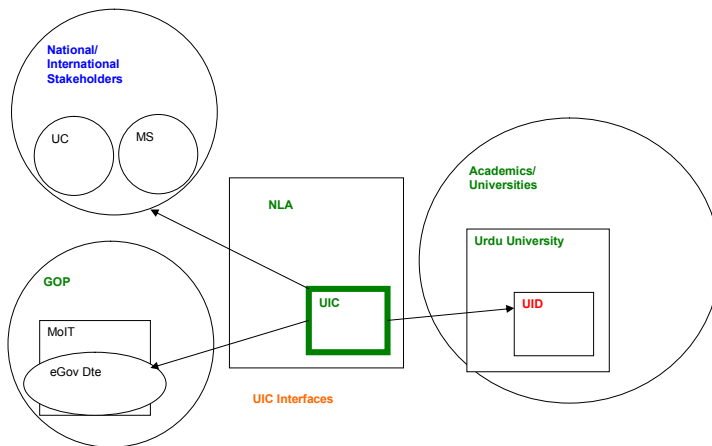
PC-I - Activities

1. Font Development- Pak Nastaleeq
2. Urdu Database– Data Development
3. Machine Translation – Official Urdu
4. OCR- Writing Pad
5. E- Learning– Prototype Training Courses
6. E- Publishing – Website and Data ware house
7. Building– Pre qualification/ designing

Targets June 2008 & 2009

1. Font Development- Pak Nastaleeq- Revision as per Unicode Provision
2. Urdu Database– Prototype DB Development and data test in September 2008.
3. Machine Translation- Prototype Official English Urdu EBMT Software Development and release in September 2008

Center of Excellence for Urdu Informatics – PC-1
PHASE I – Development of Standards
PHASE II – Establishment of Urdu Informatics Center
PHASE III – Urdu Laboratory Products/ Institute
Phase-1 -- Cost 14.586 M
Phase-2 -- Cost 37.121 M
Phase-3 -- Cost 280.00 M

مرکز فضیلت برائے اردو اطلاعیات
National/ International Stakeholders
UC
MS
NLA
Academics/ Universities
Urdu University
GOP
UIC
UID
MoIT
eGov Dte
UIC Interfaces

COMPUTATIONAL LINGUISTICS?

- ***Computational linguistics (CL)*** is a discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences and overlaps with the field of ***artificial intelligence (AI)***, a branch of ***computer science*** aiming at computational models of human cognition. Computational linguistics has applied and theoretical components.

Future Needs - مستقبل کی ضروریات

- 1. R&D- A Continuous Process for Computational linguistics.
- 2. Urdu Data Bank- A Continuous Process- with an Interval of 5 years.
- *(90% Completion is expected in 1o years)*
- 3. E-Teaching/ Learning- Teaching Courses, Certification, Training.
- 4. E- Publishing- A Continuous Process.
- 5. Laboratory Product- Software/ CD's/ Programs, Tools etc.

An Institute
for
Urdu Informatics is Needed.

Machine Translation

(a) In search of Common Lingual Code.

(b) Steps in M.T.

1. Lingual Equivalents and Properties/ Diction/ Simple Sentences.
2. Clause Equivalents and Rearrangements – Basic Formula:

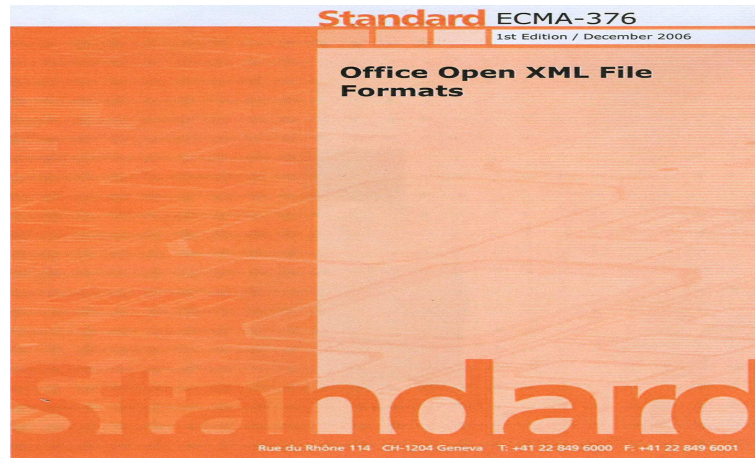
   $1/aN + n + (n-1) + (n-2) + (n-\alpha) + 1/bV$

1. Activities

   Induction, Gender, Plurality, Conjunctions etc.

Urdu Database (Software)

1. Search Items (From idea/ Meaning to word/ Phrase).
2. Filing Arrangements.
3. Synonyms and Homonyms.
4. Meaning– Relationship/ Thesauri
5. Data Development
6. Testing
7. Urdu Databank (Corpus)

ISO- Open Office Standards
Standard ECMA-376
1st Edition / December 2006
Office Open XML File Formats
Standard
Rue du Rhône 114 CH-1204 Geneva T: +41 22 849 6000 F: +41 22 849 6001

Automatic Language Translation
• Automatic language translation is the use of a computer program to translate input text from one national language to another while maintaining the original document format. Yahoo and some other sites offer what is sometimes called instant translation using such a tool. Since language is heavily dependent on context and connoted as well as denoted meaning, a program needs to have access to such context as well as the ability to use it. Since providing enough context is difficult, automatic language translation thus far seems to be successful only in limited and well-understood situations and as a first time-saving step toward translation (or "post-editing") by a human being.

Proposal to Unicode

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1].
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Preliminary Proposal to add Nuqta Characters to Arabic Block*
2. Requester's name: *Dr. Attash Durrani, Center of Excellence for Urdu Informatics (National Language Authority)*
3. Requester type (Member body/Liaison/Individual contribution): *Full Member*
4. Submission date: *29-01-2006*
5. Requester's reference (if applicable): *N/A*
6. Choose one of the following:
   This is a complete proposal: *No*
   (or) More information will be provided later: *Yes*

B. Technical – General

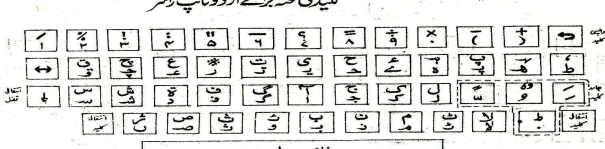
1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *No*
      Proposed name of script: *N/A*
   b. The proposal is for addition of character(s) to an existing block: *Yes*
      Name of the existing block: *Arabic*
2. Number of characters in proposal: *22*
3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary	A	B.1-Specialized (small collection)		B.2-Specialized (large collection)	
C-Major extinct		D-Attested extinct		E-Minor extinct	

- مائیکروسافٹ جیسےبین الاقوامی ادارےکو بھی اس حقیقت کا اندازہ ہو گیا تھا کہ اب لسانی توسیع میں بھی بہت سا کاروبار پوشیدہ ہے، چنانچہ دوسری زبانوں کی طرف بھی توجہ دی جانی چاہیئے یا انھوں نےاکیسویں صدی کےآغاز ہی میں اپنی تمام تر مصنوعات کو دنیا کی منتخب کردہ 100زبانوں میں پیش کرنےکا فیصلہ کیا۔ انھی 100 زبانوں میں سےایک اردو ہےجو اب عربی کےبعد ان پہلی چار زبانوں میں شامل ہےجن میں مائیکروسافٹ کی مصنوعات پیش کی جا رہی ہیں، چنانچہ کمپیوٹر سکرین کو اردو میں بدلنےکےسلسلےمیں مائیکروسافٹ اور مقتدرہ قومی زبان کےدرمیان ایک دوستانہ معاہدہ ہوا۔ جس کی رو سےمائیکروسافٹ کی مصنوعات/ اپیلی کیشن اردو میں ڈھالنےکا آغاز ہوا۔

UNICODE Ver 4.0

NLA's Keyboard adopted by
MICROSOFT WINDOWS XP
On-Screen Keyboard
File Keyboard Settings Help

Microsoft Local Language Program
"The National Language Authority of Pakistan is grateful to Microsoft for its great initiative to collaborate in the development of the Urdu Language for use in Informatics through its Local Language Program. The National Language Authority, in collaboration with Microsoft, is working to bring Computer Technology to Urdu. Providing the interface in Urdu will boost IT development activities in Pakistan, as well as in India and other SAARC countries. This initiative will potentially enable computer access in every street and village of Pakistan and South Asia. Teaching of Urdu will be facilitated and education in school can now be enhanced with the help of computers, because 92 percent of the population of Pakistan does not speak English. In addition, the software development industry will gain a new field of business activities using Office and Windows in Urdu, and we expect that the LLP will also benefit Urdu informatics research activities in Pakistan's universities."
Dr. Attash Durrani
National Language Authority of Pakistan
Press Release:
Microsoft Enables Millions More to Experience Personal Computing Through Local Language Program - March 16, 2004

THE GAZETTE OF PAKISTAN, EXTRA., DEC. 6, 1980 [PART III

III. The public sector companies concerned may undertake the manufacture of Typewriters and Teleprinters with the standardized key boards only.

Also resolved that this resolution should be published in the official Gazette.

By order of the President.
ZAHUR AZAR,
Cabinet Secretary.

ANNEX - I

کلیدی تختہ برائے اردو ٹائپ رائٹر

فاصل

مقتدرہ قومی زبان

Authorization for Standards Development

GOVERNMENT OF PAKISTAN
CABINET SECRETARIAT
(CABINET DIVISION)

SUBJECT: CABINET DECISION IN CASE NO.138/17/2000 (II), DATED 23.08.2000 – USE OF URDU FOR COMPUTER APPLICATIONS

The Cabinet in its meeting held on 23rd August 2000, decided, inter alia, as under:

"The Cabinet decided that the development of standards for the use of Urdu for computer applications shall continue to be the responsibility of the Cabinet Division."

2. The implementation status of above decision may please be furnished to this Section.

(Imam Bux Bhatti)
Section Officer

Cabinet Division, (Mr. Muhammad Ashraf, Deputy Secretary), Islamabad.
Cab. Divn. U. O. Note No.89/Prog/2000, dated 4th December 2000.

Prime Ministers Directive

Prime Minister's Secretariat (Public)
Islamabad

Subject:- MINUTES OF MEETING HELD TO REVIEW GOALS/TARGETS OF THE CABINET DIVISION ON 11-02-2006.

The Goals/Targets pertaining to the Cabinet Division were reviewed by the Prime Minister in a meeting held on 11-02-2006 during which the Prime Minister directed as follows:-

(viii) The promotion of Urdu language by making it a computer language be put on fast track

Vision 2030
Government of Pakistan
Working Draft
Vision 2030
Vision 2030 Committee
Planning Commission
Government of Pakistan
Islamabad
9th March 2007

"Urdu will be made
the language of internet"
• True sustainability, however, will come when these languages create their synergies with global modern movements and ideas, especially the Internet.
• The Case for Urdu: Urdu is and must remain the first language of Pakistan. It is and must remain the language of our culture and of our day-to-day communication. It is and must remain the first language a Pakistani child learns to speak and write. Urdu as the medium of instruction for literary subjects and general studies is a natural consequence of Urdu's position as the first language.
• Imagine a Pakistan where Urdu is taught as a second language (as is meant for foreigners). We would be in the same unhappy position as a brotherly country which holds one of the most extensive administrative and cultural treasures of written records in one of its old palaces, but its citizens cannot read them because the script was changed to Roman.

Advantages to the Society

- Urdu Computing platforms will attract masses to these platforms hence eliminating the “digital divide” in our society.
- Hand held Urdu devices will transform our society into a more interconnected society having uncountable benefits.
- Automatic Machine Translation from English to Urdu will enable the masses to access the data that, before translation, is foreign to them.
- When it comes to Scientific Text, Automatic Machine Translation will help to translate science and technology literature from abroad into Urdu thus raising the overall intellect of the society.
- Electronic Urdu Database will help to standardize the language among masses thus boosting the research and refinement of the language which definitely has a positive impact on society.

Advantages to the Economy

- Urdu Computers will encourage less literate businessmen to track their business with digital devices
- As a result of the previously mentioned consequence, Enabling Urdu platforms will boost the local Software industry to produce Urdu Business Solutions.
- Small business will start to track their business with Urdu enabled devices which in turn will generate both, the business for software industry and revenue for the central government.
- Urdu platforms will force the Hi-Tech companies to release their products specifically aimed at Urdu consumer market.

Vision Paper(2)

- ***Microsoft Urdu Office 2003*** is released by the Microsoft ***Microsoft Urdu Windows XP*** released.
- **BENEFITS:**
- This initiative will potentially enable computer access in every street and village of Pakistan and South Asia. Teaching of Urdu will be facilitated and education in schools can now be enhanced with the help of computers, because 92 percent of the population of Pakistan does not speak or use English.
- The future projects such as Urdu Databases/ Data Bank and for all their computer needs, without any extra expenditure for an Urdu support to computing, will also be a support to all others working in computing.
- Before intervention of NLA, Unicode (the International Standard for Computers) was adopting Urdu characters and standards erroneously from Arabic code plate and other unreliable sources, specially from India. With NLA now as a Full Corporate Consortium Member of Unicode for last six years, have been effectively warding off interference in Urdu language.

Advantages to the State

- Having everything in Urdu on Computers will promote national integrity.
- E-Government can take advantage of Urdu Informatics at large by having everything in Urdu accessible for masses.
- Hand held Urdu devices will help to cover the state operations to less literate areas where English isn't known.
- Will add up to a strong localized electronic infrastructure equally accessible from all parts of the state.
- Automatic Machine Translation will assist the state to keep an eye own world opinions in their own language by continuously translating from various sources.

مرکز فضیلت برائے اردو اطلاعیات

9

Institute
for
Urdu Informatics

January 15, 2008

Dr. Attash Durrani

Vision Paper

- By the end of 20th century, it was universally agreed by the linguists and computer technologists that only those languages would survive in the 21st century, which can be effectively used as a Computer Language.
- The computer giants like Microsoft, realizing the fact that computer usage can no longer be restricted to Word Processing but instead be expanded to create a database for research and development and internet usage, decided to convert and produce its soft wares in one hundred different languages. Urdu is one of these one hundred languages.
- Urdu Informatics Department of NLA devised so many projects and standardized many tables which are now being used by the computer world, NADRA, Google, Windows XP, Nokia, IBM etc. Changing Computer screen from English-to-Urdu is making it usable for an ordinary Pakistani to operate it easily.

Microsoft
Pakistan Developer Conference '06
Microsoft
Your potential. Our passion.
So, Why a Tool is Needed
• Tool = Standard Glossary
"Electronic Dictionary
of Localization of Computer Applications"
(English-Urdu)
• Guide lines:
Microsoft Urdu Style Guide
• Published by:
NATIONAL LANGUAGE AUTHORITY,
PAKISTAN

Microsoft Pakistan Developer Conference '06 — Microsoft. Your potential. Our passion.

LSP for Computer (Synonyms With Urdu Translation)-7

English	Lexical meanings	Convened by NLA
Customize	مخصوص بنانا	اختصاص کرنا
Specify	مخصوص بنانا	تخصیص کرنا
Choose	منتخب کرنا	انتخاب کرنا
Select	منتخب کرنا	منتخب کرنا
Instrument	اوزار	آلہ
Tool	اوزار	ٹول
Device	اوزار	آلہ / ڈیوائس
Fixing	درست کرنا	درست کرنا
Correcting	درست کرنا	تصحیح کرنا
Rectifying	درست کرنا	صحیح کرنا

Microsoft Pakistan Developer Conference '06
June 28-30, Karachi & June 28, Lahore

Microsoft
Your potential. Our passion.

LSP for Computer (Synonyms With Urdu Translation)-5

English	Lexical meanings	Convened by NLA
Tasks	کام	مجموعہ کار
Advance	اعلی	ایڈوانس/ اعلی
High	اعلی	اعلی
Alert	ہشیار	انتباہ
Warning	تنبیہ	تنبیہ
Allow	اجازت دینا	اجازت دینا
Permit	اجازت دینا	اجازت دینا
Action	فعل	عمل
Verb	فعل	فعل
Apply	جوڑیں	لگائیں

Microsoft Pakistan Developer Conference '06
June 28-30, Karachi & June 28, Lahore

Microsoft
Your potential. Our passion.

LSP for Computer (Synonyms With Urdu Translation)-6

English	Lexical meanings	Convened by NLA
Paste	جوڑنا/ لگانا	جوڑنا
Connect	جوڑنا	جڑنا
Complete	مکمل کرنا	مکمل کرنا
Finish	ختم کرنا	تکمیل کرنا
Over	مکمل/ تمام	ختم
Confirm	تصدیق کرنا	توثیق کرنا
Certified	تصدیق شدہ	تصدیق شدہ
Authorised	تصدیق شدہ	مجاز
Verified	تصدیق شدہ	توثیق شدہ
Custom	مخصوص کرنا	مخصوص

Microsoft Pakistan Developer Conference '06
June 28–30, Karachi & June 28, Lahore

Microsoft
Your potential. Our passion.™

LSP for Computer (Synonyms With Urdu Translation)-3

English	Lexical meanings	Convened by NLA
On	پر	بر
Bottom	پیندا	پایاں
Down	نیچے	نیچے / پست
Under	نیچے	تحت
Below	نیچے	نیچے
Beneath	نیچے	نیچے
Change	بدلنا	تبدیل کرنا
Replace	بدلنا	بدلیں
Convert	بدلنا	ادل بدل کرنا
Correct	صحیح کرنا	صحیح کرنا

Microsoft Pakistan Developer Conference '06
June 28–30, Karachi & June 28, Lahore

Microsoft
Your potential. Our passion.™

LSP for Computer (Synonyms With Urdu Translation)-4

English	Lexical meanings	Convened by NLA
Rectify	صحیح کرنا	تصحیح کرنا
Fix	صحیح کرنا	درست کرنا
Adjust	صحیح کرنا	بمطابق بنانا
Exact	صحیح	درست
Precise	صحیح	عین
Back	پیچھے	واپس
Backward	پیچھے کو/ پسماندہ	پیچھے
Behind	پیچھے	عقب
Work	کام	کار
Task	کام	کام

Microsoft Pakistan Developer Conference '06

Microsoft
Your potential. Our passion.

LSP for Computer (Synonyms With Urdu Translation)

English	Lexical meanings	Convened by NLA
Show	دکھانا	نمایش کرنا
Display	دکھانا	دکھانا
Presentation	اظہار/ پیشکش	پیش کش
Exhibition	نمایش	نمایش
Find	تلاش کرنا	ڈھونڈنا
Search	تلاش کرنا	تلاش کرنا
Explore	تلاش کرنا	ایکسپلور کرنا
Surfing	تلاش کرنا	سرفنگ کرنا
Location	جگہ	محل وقوع
Place	جگہ	جگہ

Microsoft Pakistan Developer Conference '06

Microsoft
Your potential. Our passion.

LSP for Computer (Synonyms With Urdu Translation)-2

English	Lexical meanings	Convened by NLA
Show	دکھانا	نمایش کرنا
Display	دکھانا	دکھانا
Presentation	اظہار/ پیشکش	پیش کش
Exhibition	نمایش	نمایش
Find	تلاش کرنا	ڈھونڈنا
Search	تلاش کرنا	تلاش کرنا
Explore	تلاش کرنا	ایکسپلور کرنا
Surfing	تلاش کرنا	سرفنگ کرنا
Location	جگہ	محل وقوع
Place	جگہ	جگہ

LSP for Computer (Synonyms)-3

- Apply, Paste, Connect
- Complete, Finish, Over
- Confirmed, Certified, Authorized, Verified
- Custom, Customize, Specify
- Choose, Select, Elect
- Instrument, Tool, Device
- Fixing, Correcting, Rectifying

Microsoft Pakistan Developer Conference '06

Microsoft Your potential. Our passion.

LSP for Computer (by Convention)

(The alternate Urdu Words should not be literary or lexical rather specific)

Word	Lexical Meanings	Convened Meanings
By default	از نقص/ از روئے خرابی	طے شدہ
Archive	در محراب/ در طاقچہ	محفوظہ
Arguments	دلائل/ استدلال	براہین
Bridge	پل	برج
Character	حرف، خوبی، کردار	کریکٹر
Child	بچہ	تابع
Animation	حیوان داری	حرکت اندازی
Browsing	چارا کھانا	براؤز کرنا
Surfing	بر موج ہونا	سرفنگ

LSP for Computer (Synonyms)

- Show, Display, Present, Exhibit.
- Find, Search, Explore, Surfing,
- Location, Place, Space, Memory
- Point, Dot, Period
- Event, Situation
- Top, Over, Upper, Above, On
- Bottom, Down, Under, Below, Beneath

Microsoft
Your potential. Our passion.

LSP for Computer (Synonyms)-2

- Change, Replace, Convert,
- Correct, Rectify, Adjust,
- Exact, Precise
- Back, Backward, Behind
- Job, Task, Tasks, Work
- Advance, High
- Alert, Warning
- Allow, Permit
- Action, Verb

Technical Problems of Translation/ Localization

- To avoid chaos (use standard translation tools, Glossary etc.)
- Urdu Phrase should not be more then 10% of the English Phrase.
- Replacing of tags at accurate spacing and localize every word though it is not necessary in general. (=Articles= The, a, an) example:

  The <MS Office> is friendly user program.

  یہ <MS Office> ایک صارف دوست پروگرام ہے
- Conversion of 12 tenses of English into 59 tenses of Urdu is a difficult job.

  Example: The birds would be flying.

  پرندے اڑتے ہوں گے/ پرندے اڑ رہے ہوں گے/ پرندے اڑتے چلے جا رہے ہوں گے
- Consistency regarding masculine Feminine, Singular Plural, Formal Informal Induction or declensions of words etc.

  (especially in verbs) جاتا، جاتی، جاتے، جاتیں، جاتے ہوئے، جاتے جاتے

  Next= اگلا، اگلے، اگلی، آگے
- Situation dependent translation such as:

  Next = (Adj.) اگلا
  (Adv.) آگے

  Back = (verb) واپس
  (Noun) عقب
  (Adv) پیچھے
  (Adj.) پچھلا

Microsoft
Your potential. Our passion.

Technical Problems of Translation/ Localization-2

- Clauses verses syntax. (English/ Urdu) Helium or Loc. Studio? (Urdu Syntax is difficult that the last clause of complex sentence in English is often at the 1st place of the Urdu Syntax)
- Replacing of words in Roman like Trademarks, abbreviations etc. (Problematic Applications)
- Word length maximum up-to 10 letters. Why?
- Avoiding subjectivity:

  See= دیکھو، دیکھیں، دیکھیے
  Search= تلاش کرنا، تلاش کریں، تلاش
  And= “،”، “اور”، “و”،
- In many cases Noun and Verb are the same.

  Searching = تلاش کر رہا ہے
  Searching = تلاش کرنا
  Searching = تلاش

Why Standardization?

(Computer does not interpret Ambiguity, take Context into account or Extrapolate)

- Unification, Consistency
- No Chaos, Inconsistency or Ambiguity
- Exchange of knowledge
- Universal Training
- Business Flow

(Computer only follows "Precision").

Why Convention

General Literacy Sense does not convey meaning as:

Cat = بلی

Felidae: خاندان بلی--Not Necessarily a cat.

Dog= کتا

Canidae: خاندان کتا -- Not Necessarily a dog.

Dog (Comput. Sc.) hexadecimal digit=13 decimal.

So, Convene neologs for خاندان بلی، خاندان کتا etc.

Translation/ Localization Process (Needs Objectivity and Precision)

- Individual vs Organization
  (Subjectivity) (Objectivity)
- Puritans vs adaptationists
  (Perso-Arabic) (Transliteration)
- Eclectic approach

(Local, Perso-Arabic, Transliterational words)

Technical Translation Features

- Technical language should based on neither literary nor journalistic discourse, rather a specific language.
- Literary discourse usage means:-
  one word = multiple meanings
- Technical language means:-
  one word = one meaning.

Microsoft Pakistan Developer Conference '06

Microsoft Your potential. Our passion.

Knowledge: Conceptology-2

7.Specific Terminology (Basic)	.3 million
8. Total English Terminology	560,000
9. Total Abbreviations, Acronyms, Initialisms	110,000
10. Neologisms (yearly) New coinage	2000 terms
11. Urdu (Trems)	.2 million
12. Urdu Words	32000

Rectification

- *Urdu is a regular language developed by the civil society of that era.*
- *Urdu is a subjective language not easily compatible to the other languages of the World i.e. English, French, Arabic etc.*
- *Easy language means journalistic/ informatic language but the technical language or discourse (LSP) has its own features.*
- *Journalistic discourse means word is equal to the general meanings.*

Knowledge: Conceptology

1.Total Sciences	250
2.Total Concepts	17 millions
3.Maximum words in any language	60,000
4.Literary discourse (Shakespeare used 22000 words)	15000 to 20000 words
5. Journalistic discourse	10,000
6. Daily Chat usage	1500 words

Microsoft
Your potential. Our passion.
Microsoft
Pakistan Developer Conference '06
June 28~30, Karachi & June 28, Lahore
8
Localization and Translation Challanges
By:
Dr. Attash Durrani

Microsoft
Pakistan Developer Conference '06
Microsoft
Your potential. Our passion.
Some wrong notions about
Urdu Localization/ Translation
• Urdu is a Pidgin (لشکری) Language.
• Urdu has already some capacities to build for and communicate every thing.
• Translation must be easily/ directly understandable.

LSP for Computer
(Synonyms With Urdu Translation)-7

English	Lexical meanings	Convened by NLA
Customize	مخصوص بنانا	اختصاص کرنا
Specify	مخصوص بنانا	تخصیص کرنا
Choose	منتخب کرنا	انتخاب کرنا
Select	منتخب کرنا	منتخب کرنا
Instrument	اوزار	آلہ
Tool	اوزار	ٹول
Device	اوزار	آلہ
Fixing	درست کرنا	درست کرنا
Correcting	درست کرنا	تصحیح کرنا
Rectifying	درست کرنا	صحیح کرنا

So, Why a Tool is Needed

Tool = Standard Glossary

"Electronic Dictionary
of Localization of
Computer Applications"
(English-Urdu)

Published by:

NATIONAL LANGUAGE AUTHORITY, PAKISTAN

Rs.= 200 **Pages = 232**

LSP for Computer
(Synonyms With Urdu Translation)-5

English	Lexical meanings	Convened by NLA
Tasks	کام	مجموعہ کار
Advance	اعلی	ایڈوانس/ اعلی
High	اعلی	اعلی
Alert	ہشیار	انتباہ
Warning	تنبیہ	تنبیہ
Allow	اجازت دینا	اجازت دینا
Permit	اجازت دینا	اجازت دینا
Action	فعل	عمل
Verb	فعل	فعل
Apply	جوڑیں	لگائیں

LSP for Computer
(Synonyms With Urdu Translation)-6

English	Lexical meanings	Convened by NLA
Paste	جوڑنا/ لگانا	جوڑنا
Connect	جوڑنا	جڑنا
Complete	مکمل کرنا	مکمل کرنا
Finish	ختم کرنا	تکمیل کرنا
Over	مکمل/ تمام	ختم
Confirm	تصدیق کرنا	توثیق کرنا
Certified	تصدیق شدہ	تصدیق شدہ
Authorised	تصدیق شدہ	مجاز
Verified	تصدیق شدہ	توثیق شدہ
Custom	مخصوص کرنا	مخصوص

LSP for Computer
(Synonyms With Urdu Translation)-3

English	Lexical meanings	Convened by NLA
On	پر	بر
Bottom	پیندا	پایان
Down	نیچے	نیچے/ پست
Under	نیچے	تحت
Below	نیچے	نیچے
Beneath	نیچے	نیچے
Change	بدلنا	تبدیل کرنا
Replace	بدلنا	بدلیں
Convert	بدلنا	ادل بدل کرنا
Correct	صحیح کرنا	صحیح کرنا

LSP for Computer
(Synonyms With Urdu Translation)-4

English	Lexical meanings	Convened by NLA
Rectify	صحیح کرنا	تصحیح کرنا
Fix	صحیح کرنا	درست کرنا
Adjust	صحیح کرنا	بمطابق بنانا
Exact	صحیح	درست
Precise	صحیح	عین
Back	پیچھے	واپس
Backward	پیچھے کو/ پسمانده	پیچھے
Behind	پیچھے	عقب
Work	کام	کار
Task	کام	کام

LSP for Computer
(Synonyms With Urdu Translation)

English	Lexical meanings	Convened by NLA
Show	دکھانا	نمایش کرنا
Display	دکھانا	دکھانا
Presentation	اظہار/ پیشکش	پیش کش
Exhibition	نمایش	نمایش
Find	تلاش کرنا	ڈھونڈنا
Search	تلاش کرنا	تلاش کرنا
Explore	تلاش کرنا	ایکسپلور کرنا
Surfing	تلاش کرنا	سرفنگ کرنا
Location	جگہ	محل وقوع
Place	جگہ	جگہ

LSP for Computer
(Synonyms With Urdu Translation)-2

English	Lexical meanings	Convened by NLA
Space	جگہ	گنجایش
Memory	یاد داشت	میموری
Point	مقام/ نقطہ/ جگہ	نقطہ/ نکتہ/ پوائنٹ
Dot	نقطہ	ڈاٹ
Dash	چھوٹا نقطہ	ڈیش
Event	وقوعہ/ موقع	وقوعہ
Situation	صورت حال/ موقع	موقع محل
Top	اوپر	بالا
Over	اوپر/ پر	بر
Above	اوپر	پر

LSP for Computer
(Synonyms)-3

- Apply, Paste, Connect
- Complete, Finish, Over
- Confirm, Certified, Authorised, Verified
- Custom, Customize, Specify
- Choose, Select
- Instrument, Tool, Device
- Fixing, Correcting, Rectifying

LSP for Computer
(by Convention)

Word	Lexical Meanings	Convened Meanings
By default	از نقص/ از روئے خرابی	طے شدہ
Archive	در محراب/ در طاقچہ	محفوظہ
Arguments	دلائل/ استدلال	براہین
Bridge	پل	برج
Character	حرف، خوبی، کردار	کریکٹر
Child	بچہ	تابع
Animation	حیوان داری	حرکت اندازی
Browsing	چارا کھانا	براؤز کرنا
Surfing	بر موج ہونا	سرفنگ

LSP for Computer
(Synonyms)

- Show, Display, Presentation, Exhibition.
- Find, Search, Explore, Surfing,
- Location, Place, Space, Memory
- Point, Dot, Period
- Event, Situation
- Top, Over, Upper, Above, On
- Bottom, Down, Under, Below, Beneath

LSP for Computer
(Synonyms)-2

- Change, Replace, Convert,
- Correct, Rectify, Adjust,
- Exact, Precise
- Back, Backward, Behind
- Job, Task, Tasks, Work
- Advance, High
- Alert, Warning
- Allow, Permit
- Action, Verb

Technical Problems of Translation/ Localization

- To avoid chaos (use of translation tools, Glossary etc.)
- Phrase should not be more then 10% of the English Phrase.
- Replacing of tags at accurate spacing and localize every word. (=Articles= The, a, an) example:
  The <MS Office> is friendly user program.
  یہ <MS Office> ایک صارف دوست پروگرام ہے
- Conversion of 12 tenses of English into 56 tenses of Urdu.

The birds would be flying.
پرندے اڑتے ہوں گے/ پرندے اڑ رہے ہوں گے/ پرندے اڑتے چلے جا رہے ہوں گے

- Consistency regarding masculine Feminine, Singular Plural, Induction etc.

(especially in verbs) جاتا، جاتی، جاتے، جاتیں، جاتے ہوئے، جاتے جاتے
Next= اگلا، اگلے، اگلی، آگے

- Situation dependent translation such as:
  Next = (Adj.) اگلا
  (Adv.) آگے

-2Technical Problems of Translation/ Localization

- (English/ Urdu) Helium Clauses verses syntax.
- Replacing of words in Roman like Trademarks, abbreviations etc. (Problematic Applications)
- Word length maximum up-to 10 letters.
- Avoiding subjectivity:
  See=دیکھو، دیکھیں، دیکھیے
  Search= تلاش کرنا، تلاش کریں، تلاش
  And= "،" ، "اور" ، "و" ،
- In many cases Noun and Verb are the same.
  Searching = تلاش کر رہا ہے
  Searching = تلاش کرنا
  Searching = تلاش

Why Standardization?

- Unification
- No chaos, inconsistency or ambiguity
- Exchange of knowledge
- Universal Training
- Business Flow

Why Convention

Cat = بلی

Felidae: خاندان بلی--Not Necessarily a cat.

Dog= کتا

Canidae: خاندان کتا -- Not Necessarily a dog

Dog (Comput. Sc.) hexadecimal digit=13 decimal.

Translation/ Localization Process

- Individual vs Organizational
  (Subjectivity) (Objectivity)
- Puritans vs adaptationists
  (Perso-Arabic) (Transliteration)
- Eclectic approach

(Local, Perso-Arabic, Transliterational)

Technical Translation Features

- Technical language should based neither literary nor journalistic discourse.
- Literary discourse usage:-
  one word = multiple meanings
- Technical language means:-
  one word = one meaning.

Knowledge: Conceptology-2

7.Specific Terminology (Basic)	.3 million
8. Total English Terminology	560,000
9. Total Abbreviations, Acronyms, Initialisms	110,000
10. Neologisms (yearly) New coinage	2000 terms
11. Urdu (Trems)	.2 million
12. Urdu Words	32000

LSP- Language for Specific Purposes

- Specific Language is for specific use.
- LSP is for the Experts and not for the general public.
- Word coinage is by convention (بذریعہ قرارداد)
- Rights of coinage = originator (Inventor, discoverer, writer, translator, etc)

Rectification

- Urdu is a regular language developed by the civil society of that era.
- Urdu is a subjective language not easily compatible to the other languages of the World i.e. English, French, Arabic etc.
- Easy language means journalistic/ informatic language but the technical language or discourse (LSP) has its own features.
- Journalistic discourse means word is equal to the general meanings.

Knowledge: Conceptology

1.Total Sciences	250
2.Total Concepts	17 millions
3.Maximum words in any language	60,000
4.Literary discourse (Shakespeare used 22000 words)	15000 to 20000 words
5. Journalistic discourse	10,000
6. Daily Chat usage	1500 words

7

URDU LOCALIZATION ISSUES AND PROBLEMS

August 28, 2005
Microsoft/ Karachi/
IBA Golden Jubilee Seminar

Some wrong notions about Urdu Localization/ Translation

- Urdu is a Pidgin (لشکری) Language.
- Urdu has already some capacities to build for and communicate every thing.
- Translation must be easily/ directly understandable.

Training

•Specific applicability areas pertaining to the language & script processing of the specific languages may be focused.

•MS Curriculum may be included in Training

The Solution

•Ghost Characters and Dots. (22+30)

•May be minimized to (22+22)

•Ghost characters are already supported by the UNICODE.

•Proceedings for addition of dots is in process.

•Will solve the problem of data entry altogether.

Training

•More people know to drive a car, More scope for cars!

•Training of the Microsoft Localized Applications is of vital importance.

•Training specifically tailored to the community needs will broaden the scope of MS Localized Applications.

Major Issues…

• Data Entry is a major issue on all the computing devices and gadgets.

•For the localized applications to be practical in Pakistan, many keyboard layouts are required at once.

•Which is not only impossible but cumbersome!

The Real Trouble

•There are 58 characters (in Urdu) and there are 8 keys all in all on a handheld device.

•And yet to talk about the other Pakistani Languages.

•Simply cant fit all the pieces onto a single canvas.

The Future….

- Business Solutions will provide low cost alternative to the SMEs and private Entrepreneurs.
- Retailers
- Distributors
- Even small shopkeepers!

Future of Computing! The Mobile…

- Mobile phones are replacing computers day by day, as many Business professionals in Europe prefer to have a Localized Smart Phone instead of Laptop.
- Pakistan holds about 20% of the total Mobile Market, making it one of the largest Mobile Phone users market in the World.

Issues and Role of Gov’t

• Tools for Linguistic Analysis are of immense need for the Urdu academia.
• Prose analysis and text grading tools have to be developed by Authority on Languages
• NLA is already helping vendors by developing Generic and Standard products such as Keyboard, Standard Font, Glossaries and Rules to perform Local Language Computing

The Future….

Following are the areas that can be much productive.

•Business solutions on top of MS Access and MS Excel

•Linguistic Analysis Tools based on MS Word

•Specialized DTP Packages based on MS Word.

•Specialized Project Management Tools based on MS Project.

Issues and Role of Gov’t

- Language Sensitivity issues
- Government can play as a regulatory authority to Standardize Local Computing in Pakistan

Issues and Role of Gov’t

- The existing functionality in the Localized Applications is not sufficient in all cases.
- The need for extensions or specialized functionalities is required.
- This definitely requires to develop customized solutions on existing code base.

Scope! (cont’d)

Localized Microsoft Applications can bring a dynamic change in following areas of digital computing

- Word Processing/Internet Browsing/Chatting/Emailing
- Printing and Publishing
- Electronic Record Keeping
- Statistical Analysis Facilities
- Project Management & Planning
- Adverting and designing

(and above all these a wider audience to address.)

Outcomes

- Promotion of National and Regional Languages
- Due to Localized versions of the Applications, Users will feel much comfortable and it may help to increase in their productivity
- Microsoft Localized Applications can bring brand loyalty among the masses!
- Revenue Generation and Follow up for other IT/Telecom Vendors

Scope!

Possible key users of Microsoft Localized Products in Pakistan

- Government Offices (Database, Records, Word processing using Microsoft Localized Word, Excel)
- Government Institutes (Database, Records, Internet, Word processing using Microsoft Localized Word, Excel, Outlook)
- Universities and Colleges (Database, Records, Word processing, Presentations using Microsoft Localized Word, Excel and PowerPoint)
- SMEs (Small & Medium Enterprises) (Record Keeping and Word Processing using Localized version of Excel, Access and Word)

Scope! (cont'd)

Possible key users of Microsoft Localized Products in Pakistan

- Services Sector (Database, Record keeping, Account Management using Localized versions)
- Home Users (Internet, Word Processing and General Computing using Localized version of Windows Excel, Access and Word)

Why Local Lang. Computing?

<u>For Software Vendors</u>

- Target Audience for ground hitting results
- More Users
- More Revenue
- More Creditability
- Low Cost Development

Global Business of Local Computing

<u>Turn over of Local Language Computing Business in 2005</u>

Total Global Localization Business of All Localized Languages

USD 8.8 Billion

Pakistani Languages Share

USD 25 Million

Localization Business Generated/Granted to Pakistan

USD 2 Million

Source: Common Sense Advisory - USA

Local Languages in Pakistan

Local Languages

•**Punjabi:**	44% of the national total
•**Pashto :**	15%
•**Sindhi :**	14%
•**Siraiki :**	11%
•**Urdu :**	8%
•**Balochi:**	4%
•**Others :**	4%
•**English :**	**(Communication language of most organizations and ministries)**

Why Local Lang. Computing?

Because!

- Over 80% (135M out of 140M) of Pakistan's population cannot benefit from IT Revolution due to language barriers
- About 60% of India's population is ignored from IT benefits

To compete with other nations, it is the need of time!

Introduction

Pakistan: Population

- 165,803,560

Literacy

- Definition: age 15 and over can read and write.
- total population: 48.7%
- male: 61.7%
- female: 35.2%
- Only 2% of 48.7% are Anglo Literate!

The Need for Local Language Computing Standards

- Informatics and Localization are relatively a new field in developing countries, it is the need of the day to inform people and create awareness regarding the importance of Local Language Computing and Achieving Digital Literacy with Localization.
- Most of Computer Standards need to be “Re-born” for Local Language Needs, which can be done by arranging workshops, seminars to discuss the Language Needs and Computer Standards with Language and IT experts.
- New Standards such as Microsoft Office XML can play an important and vital in achieving digital literacy in Pakistan by providing support to Local Languages, as currently none of other software vendors are providing such support.

The Future

1. The other projects in CEUI are Optical character recognition Software with XML support it can be made further possible to put tags among Characters and Words to Recognize, which may bring another bench-mark in Multilingual Document reading softwares.
2. Knowledge and Example based Machine Translation Software.
3. Urdu Data Bank for corpus of 1Billion words.
4. Other research and teaching in Urdu Informatics, Computational Linguistics, Multilingual Artificial Intelligence softwares etc.
5. Curriculum Development.

UNICODE Support

As full corporate member of UNICODE Inc. USA, we tried to get full support of UNICODE, ISO etc.

Based on our recommendations, initially partial support was given in Unicode Ver. 3.1 for a s basic ghost character set and doubting for the Nuqta proposal.

By March 2008, the technical committee of UNICODE agreed with our proposals and they allotted Unicode Support from 0880 code in Unicode. (Approved by UTC on 11-12 May 2008).

The developed Urdu Font is fully compliant with any standard and it can easily be used with upcoming Microsoft applications such Microsoft Vista, Windows Mobile 7 and Office 2007, etc.

Unicode Support for Nuqtas

0880 ARABIC SINGLE NUQTA ABOVE
0881 ARABIC SINGLE NUQTA BELOW
0882 ARABIC DOUBLE NUQTA ABOVE
0883 ARABIC DOUBLE NUQTA BELOW
0884 ARABIC TRIPLE NUQTA ABOVE
0885 ARABIC TRIPLE NUQTA BELOW
0886 ARABIC TRIPLE INVERTED NUQTA ABOVE
0887 ARABIC TRIPLE INVERTED NUQTA BELOW
0888 ARABIC QUADRUPLE NUQTA ABOVE * Sindhi
0889 ARABIC QUADRUPLE NUQTA BELOW* Sindhi
088A ARABIC DOUBLE DANDA ABOVE * Sindhi
088B ARABIC DOUBLE DANDA BELOW* Sindhi
088C ARABIC DOUBLE NUQTA VERTICAL ABOVE * Sindhi
088D ARABIC DOUBLE NUQTA VERTICAL BELOW * Sindhi
088E ARABIC SINGLE KASHIDA ABOVE * Urdu
088F ARABIC SINGLE KASHIDA BELOW * Urdu
0890 ARABIC DOUBLE KASHIDA ABOVE * Urdu
0891 ARABIC DOUBLE KASHIDA BELOW * Urdu
0892 ARABIC SINGLE CIRCLE ABOVE * Pashto
0893 ARABIC SINGLE CIRCLE BELOW * Pashto
0894 ARABIC TOTA ABOVE * Urdu
0895 ARABIC TOTA BELOW * Urdu

Ghost Characters Theory

At CEUI, we are not just translating the Computer applications, but also working on a single universal Urdu Characters font for all complex Arabic Script Languages.

Using the latest cutting edge technologies, to bring singularity in all the Arabic based languages for all their now and future characters in Desktop and Mobile Computing needs such as Windows Operating Systems for Both Desktop and Mobile users (for. e.g. Windows Mobile, Windows Vista etc.)

How It is possible?

Only with a new basis i.e. Font Ghost Characters Theory: Only 44S ghost characters can do this for all Arabic based language and complex script. It is ideal for both Desktop and Mobile Computing and Word-Processing needs using XML it can be implemented easily.

As it requires less disk and RAM space and equally good for Mobile Operating systems running on slow processing power.

Rationale

Arabic script was historically a "dot less" script. By this we mean that a single shape may have different sounds depending on the word. Here is an example

Fig1: Arabic phrase without dots. Native speaker has no difficulty

In the figure above, a native Arabic speaker is able to comprehend the meanings of text based on context and his/her vocabulary. However, anyone less familiar with Arabic language will not be able to understand the correct meanings of the text because of limited vocabulary and unable to understand the context. The main reason of not being able to read such a text is that the sound of a character is heavily dependent upon the context and content of the text.
To overcome this problem, Nuqta were introduced by a Muslim Caliph. The sole purpose of the dot was to sit on a shape (which we call basic shape or Kashti) and to depict its phonetic status. Below is the "dotted" version of the above mentioned text.

Fig 2: Arabic phrase with dots. Sounds of characters are not to be "guessed"

Now after the placement of dots, even a non-native reader can easily understand the text without any hit and trail because dots are sufficiently depicting the exact sound of the character.
Later on, when new languages adopted Arabic script as there script of choice, a new problem arose and that was un-availability of sounds. For example, the Urdu has a sound exactly equal to sound of "p" in English but Arabic language has no such sound and hence script has no means to depict this sound. Again Nuqta to rescue, this problem was solved by taking the basic shape of bay and placing three dots beneath it. Here how it looks

Major Urdu Localization Completed Projects

- Urdu Localization of Windows XP, Office 2003 performed by National Language Authority, for Microsoft.
- Urdu Localization of Windows Starter Edition performed by INKSOFT.
- Urdu Localization of Windows Vista and Office 2007 translated by FAST NU, and validated & reviewed by INKSOFT.
- Urdu Support Developed for NOKIA, Motorola, and Samsung Mobiles by INKSOFT.
- Urdu Text Editor and Support for Windows Mobile 6.1 by INKSOFT
- The Standard Terms Translations (more than 5 Million Words used in Microsoft applications) have already developed and approved by NLA.

Urdu Informatics

"Urdu Informatics" is relatively a new area established in IT field.

As it is quoted in Wikipedia, the web based encyclopedia, "Urdu Informatics (اردو اطلاعیات) relates to the cutting-edge research and efforts in bringing the utilities and usage of National Language to the modern information and communication technologies in education and businesses. Works in Urdu Informatics also relate to the use of Urdu for internet applications, input controls, dynamic association of keyboard buttons to Urdu alphabets and use of proper character coding schemes."

Global Localization Business

- Localization is a Global Business.

  Total IT based Localization Business in 2005 was US$ 8.8 Billion, which has now increased to US$10 Billion.

  (Urdu is one of the four Top most Microsoft Localization Languages. Out of this business Urdu Localization Business is US$ 25 million only)

Source: Common Sense Advisory, USA.

Digital Literacy in Pakistan

To achieve our goal of eliminating "Digital Illiteracy" and Achieving Digital Inclusion in Pakistan.

Localization in Local Languages is the main key to shorten the Technology Gap and outcomes of creating technology rather than working can only be achieved with Localization. Otherwise we are limiting the knowledge to 0.02% people of Pakistan.

Localization is the only tool for "Digital Divide" in our Society. The society can be inter-connected having uncountable benefits using Localized masses.

Most Successful Example of Localization

- Homeopathy, a German Science of Treatment, but more indigenous in India and Pakistan.

  Homeopath practitioner is found in every village and street of the South Asia, because Homeopathic Courses are taught, learnt, and practiced in Urdu and Local other Languages of South Asia rather than in English or German and all drugs of Schwabe Germany are sold here with their German nomenclature.

Outcome: Homeopathic Science is accessible to millions due to Localization (Urdu).

Localization is taught as a Discipline

- Localization is not only a general local issue but also studied at Postgraduate, Masters even at Ph.D level as a discipline etc.

  The Localization Research Centre, Dublin is the information educational and research centre for the localization community, established in 1995 at University College Dublin under the Irish Government and European Union funded Technologies Centers Programme.

  Localization (Education & Research) facilities are now spreading such as:

- Postgraduate Localization Courses and Masters Degree in Global Computing, at the University of Limerick
- Masters Degree in Game Localization, funded by EA USA.
- Quality Assurance in Digital Content Localization PhD Courses.
- STAR Services Linguistics, Localization Certification professional courses.
- SDL Localization Certification professional courses, London UK

Need for Localization?

About 45% knowledge of the world exists in English whereas remaining 55% is in other European languages like:

- Russian,
- French, and
- German etc.

Since the world is becoming a global village, it is realized internationally that if the products and their uses is narrated in the respective language of a nation then the implementation and execution of the said product can be enhanced many times.

Need for Localization?

Pakistan with an 37% literacy rate, out of which only 2% can read and understand English.

Rest of 35% are out of bound for the digital world. As 35% are unable to understand English or communicate their views in English.

How we can pass the bay of digital divide and eliminate "Digital illiteracy"?

Answer: Localization of Technology (and Curriculum) on such environment which supports Local Language Computing fully i.e. XML and development of standards etc.

What is Localization?

- Localization (short: "L10n") is the process of adapting and converting any product or service into a particular language and culture, to reach non-English speaking market.

  It can be a translation of computer software, website, manual guide of product, subtitling of Video, voice-overs, computer games, Textbooks, Demos, Weights and measures, Names and titles into specific and local culture.

Need for Localization?

Mark Davis of UNICODE Inc. on January 22, 2003 described as GDP by language

"Many people in the software industry don't realize how important it is to localize products for different languages around the world. While English is a major language, it only accounts for around 30% of the world Gross Domestic Product (GDP), and is likely to account for less in the future.

Commercially, it is significant to internalize the Technology products by Localizing it into various languages.

5

Localization as a means to Achieve Digital Inclusion

Dr. Attash Durrani
Digital Inclusion Roundtable at Microsoft Events
Karachi, March 11, 2008 and Islamabad March 12, 2008.

Change in Digital Divide Rates

The digital revolution rapidly developed and now digital divide is clearly identifiable in every aspect of life.
The main reasons which are widening this gap day by day, are:

1. Limit to access Knowledge
2. Non-availability of knowledge in Local Languages
3. Lack of ability to produce knowledge

If proper concentration will be given this gap is likely increas e and we may face the issue of "Digital illiteracy" in developing countries.

This digital divide of society could only be eliminated with the help of:

1. Anglicizing the Society
2. Localization of "Technology" and "Curriculum" into Local Languages.

issues and their solutions: i.e. take for Urdu; equally admissible for other Pakistan languages.

(Abstract of Presentation made in Shaikh Ayaz International Conference on Language and Literature, Faculty of Arts, University of Sindh, Jamshoro, January 10, 2007).

4

Localization of Computing:
Lingual Issues and Solutions

Localization of computing for Urdu and other Pakistani languages is a daunting task and may be led on to wrong dimensions due to the wrong assumptions that Urdu is a Pidgin language and the translation process will be very obvious. Rather the fact is that Urdu (including other Pakistani languages) is not a Pidgin language. it's a regular language developed by the civil society of that time and due to its subjectivity, it can not be compared with English, Arabic and French. Moreover, when it comes to localization of computing, the usage of general or journalistic language may lead to ambiguity. Just to have a view on the complexities, we have 250 sciences containing 17 million concepts. On the other hand, to express these concepts, any language can have maximum 60,000 words. Literature uses somewhat 15,000 to 20,000 words and it was the Shakespeare who alone used 22,000 words. For day to day conversation 1500 words are more than enough and for journalistic usage 10,000 suffice. Now, if we deploy the journalistic discourse to depict the technical meanings, we are mapping 10,000 entities to 17 million which will definitely yield ambiguity.

Therefore, for the computational aspects, journalistic or general discourse is not adequate rather a specific dialect or Jorgan is needed and same is the case with other Pakistani languages. Everywhere we have to look-into a LSP (Language for Special Purpose). The paper/presentation stands for lingual

code plate for Urdu last year. In addition, some progress has also been made to incorporate it under the Unicode scheme, but so far there has been no concerted effort to set up domain names, email and Web addresses in Urdu. Without putting such a system in place, true and meaningful Urdu Internetworking would not be possible.

the quintessential symbol of our times. It is the harbinger of a globalized information age, which is going to replace the products of industrial manufacturing with specialized knowledge - packaged as information bytes - as the highest commodity.

But are we ready in Pakistan to embrace the Internet and actualize the great potential of this universal tool? Recently there has been a proliferation of the Internet in Pakistan, its growth fuelled by the Universal Internet Access scheme of the government, coupled with drastic reduction in bandwidth rates. No longer confined to major cities alone, the Internet is now accessible from hundreds of small cities and towns all over the country. However, it's still being used by a small minority that can read and write English.

What does the Internet offer to a Pakistani, who is part of a vast majority that doesn't know English? One hardly needs any complex search techniques to realize the paucity of Pakistani content on the Web. One of the principal reasons for this unfortunate state of affairs is the difficulty of using Urdu - and other Pakistani languages - on the Internet. It is also exacerbating the already stark 'digital divide' - between the haves and have-nots - to between those who are well versed in English and those who are not.

The Sustainable Development Networking Programme (SDNP), a global programme of UNDP, managed by IUCN - The World Conservation Union, in Pakistan, has been trying to address this question. During the last few months, it has organized a number of training workshops in Urdu and Sindhi Web publishing, in addition to a variety of other initiatives like the development of district websites and Cyber Community Centres (CCCs) in remote areas of the country. Urdu Web publishing has been facilitated by the adoption of a standard

Urdu letters. This sub-committee also has to develop a Transliteration Dictionary for the use of vendors, software houses and ISP's. The ASCII Urdu Code page Ver-2 and the Keyboard (Ver-1) has been notified by the Cabinet Division. The time has now come to consolidate all individual efforts into one platform of standardization for the betterment of the future of Urdu in IT.

(Read in a Workshop MINC on 28 July, 2001 at Islamabad)

(News) The Workshop

In collaboration with the Multilingual Internet Names Consortium (MINC) - a non-profit, non-governmental, international organization - and with the support of the National Language Authority (NLA), Pakistan, SDNP was holding an international workshop on Urdu Internet domain names, keywords and routing at Islamabad on July 28. This workshop would provide a forum for national experts in Urdu informatics to deliberate on the issues of developing an Urdu Internet names system and its applications. A working group would be formed at the end of the workshop which will carry forward this work and liaise with MINC in future. Apart from the CEO of MINC, representatives of the Arabic Internet Names Consortium are also expected to participate in this event.

The Urdu Internet Working Group, to be formed as a result of this workshop, under the auspices of the National Language Authority, would go a long way in removing the technical impediments in the Use of Urdu on Internet.

Background

Providing a smorgasbord of multi-media offerings, ranging from information and knowledge to pure entertainment, and from art to commerce and trade, the Internet has become

language - only the script is different because it is written in the Devanagri.

The Nastaleeq script is complicated because its characters show different shapes and places in different ligatures. For example letter "BAY" gives 35 shapes and placements. Approximately 6500 places are required to encapsulate the various shapes of Urdu onto the Urdu Zaabta Takhtee (UZT) for their internal representation along with Airabs and symbols etc. Also, the Nastaleeq ligature technically does not proceed from right to left. It is rather left to right for a ligature. The last and full shape of a character determines the shape and place of previous character so the ligature falls from right to left.

Mr. Ahmad Jameel Mirza decided to give ligatures as a base in the memory. This was a successful story as long as the word processing and desktop publishing is concerned. But if you require letter-based software, if you need to use Urdu Nastaleeq for email and websites using a long memory for images required in ligatures, the ligatures-based system fails. Email and websites are letter-based; whereas the Unicode is concerned only with the script and font, not the language. Its Arabic page gives additional characters, Tay, Dal, Rray, Pay, Chay etc. for Persian, Malay, Morocco and Urdu usage but gives two types of Naskh script, and too many ligatures for their combinations in these two Naskh fonts. It is least bothered about the Urdu Nastaleeq script used in Pakistan.

On the home front, a sub-committee of NLA on Urdu Email/Internet, is working under the coordinatorship of Mr. Zaheer Aziz from International Islamic University. They have evolved a standard for Reversible Roman Transliteration of

3

Linguistic Aspects of Urdu Informatics Standardization

Summary of Presentation in MINC

We have been dealing with the issue of computing in Urdu since 1984. After all these years, however, the question remains as to why Urdu could not achieve its real place in the realm of Computer Science and Information Technology.

Although individual Urdu scholars, students and writers have made concerted efforts in getting their scripts composed i.e. word processed, the problem is one of uniform standardization of the process. This needs to be maintained on all fronts, be they of a technical or a linguistic nature.

It is a linguist's responsibility to take into account the various difficulties facing him while standardizing an Urdu code page and keyboard. The complexities of the Urdu Nastaleeq script in use in Pakistan have to be understood and accommodated by the technical experts in finalizing the finished product. In this context, a linguist's viewpoint is of critical importance.

Naskh and Nastaleeq are two different scripts having their own set of fonts. Urdu is not a subset of the Arabic language. It is an Indo-Aryan language, having assimilated many words from Arabic and Persian. Urdu in Pakistan is basically written in the Nastaleeq script. In India - for the Hindi

and strive towards a peaceful and prosperous future. "To improve the quality life for all in the region, it is essential for governments to substantially increase funds to social sectors by reducing expenditure on heavy armament.

"South Asia should be made a nuclear free zone. Without cooperation, interaction and mutual support it is not possible to address and resolve socio-economic issues of poverty, unemployment, gender inequality and illiteracy. The peace efforts of the governments of India and Pakistan are welcomed," said the declaration.

The declaration said that cooperation among writers and intellectuals was essential to ensure the literature of the SAARC region translated into other languages.

The declaration urged the governments of SAARC countries to protect human rights, especially of women and underprivileged people. "Free and low cost trade of books, print material and electronic media programmes are essential to exchange ideas and writers should be given visa exemption or multiple visa for travel within the region."

There were several recommendations at the concluding session of the conference that suggested that reaching out to the widest cross section of society was imperative to enhance the vibrancy of creative writing and dialogue. The conference stressed the need for more cooperation and collaborations through the SAARC writers organization.

The conference participants walked together on The Mall as a token peace march at the end of the day. They held banners and placards inscriber with slogans for peace and tolerance. The delegates will leave for their countries Tuesday.

The 11th SAARC Writers Conference will be held in New Delhi in September 2005.

She also protested at the commercialization of embassies and visa fees for travel to other SAARC countries like Nepal and Sri Lanka.

She thanked the participants, the organisers and others for their help in making the conference a success. She also thanked Federal Education Minister Zubaida Jalal for her cooperation.

Ms Cour congratulated the Pakistan chapter of the forum for arranging the conference in Lahore. She urged writers to do more for peace and literature in the region.

Action Aid's Dr Fouzia Saeed urged SAARC countries to strengthen their ties for a peaceful South Asia.

There were three sessions on the concluding day of the conference. The first was 'The importance of oral Lliterature in classic and contemporary literature and its keynote speaker was Intizar Hussain (Pakistan). Khadim Hussain (Pakistan) and Professor Afzal Hussain (India) chaired the second session titled the 'Emergence of globalization and its reflection in literature'. The third session, 'Linguistics, Information Technology and Lliterature', was chaired by Dr Attash Durrani (Pakistan).

Mustansar Hussain Tarar read the Lahore Declaration and recommendations at the end of the concluding session.

Peoples of this region desire peace and cooperation to ensure that their basic needs and rights are fulfilled. Uterature and the arts in this region reflect this desire in multiple ways," according to the declaration.

It said that a continuous process of exchange and publications of creative and analytical writings for the SAARC region will enable all to share concerns, socio-political realities

we have already taken, an unprecedented wave of support for improved relationships has become evident in both countries."

Mr. Chandra said both countries were at the start of a new and important phase in their relations. "We must continue in the direction that we have started moving towards in the last few months, responding to popular expectation and the enthusiasm of the people - doctors, parliamentarians, journalists, artists, judges, writers, scholars, fashion designers and ordinary people - who have established links across the border," he said, adding he believed the writers conference was a step for further people-to-people contacts in South Asia.

"We have come to realise in India over the years the important and crucial role that a civil society can play in contributing to policy formation," he said. He lauded the efforts of Ajeet Cour and Kishwar Naheed for organising the conference.

Ms. Naheed also spoke about Indo-Pakistan relations, because it was easier to get visas to other SAARC countries than India. She demanded that the role of intelligence and security agencies in clearing people wishing to travel to India be ended.

"No improvement in relations and people-to-people contact can happen until the end of these unfair checks on both sides," she said. She apologised to the Indian delegates on the "misbehaviour" of the Pakistan Rangers when they arrived at the Wagah border crossing.

Ms Naheed said they would have to change their mindset for better relations, otherwise all efforts were useless. "Why do we have to wait for an invitation or a death or marriage of a relative to travel?" she said.

Daily Times, Monday, March 15, 2004

SAARC Writers Conference March 12-14 LAHORE: by Waqar Gilani

The conference, titled "Understanding and Synergy Through Literature" concluded with a declaration that called for a Inuclear-free South Asia and urged member countries to protect human rights, especially the rights of women.

The conference, held at the Alhamara Hall from March 12 to 14, aimed to promote peace and literature through the process of understanding. It was organised by the Ministry of Education, Hawwa Associates, Action Aid Pakistan and the Pakistan Academy of Letters.

Dr. Namwar Singh, the chairman of the concluding session, lauded the efforts of the Pakistan chapter of SAARC Writers for arranging the first conference in the country. He said copyrights should be enforced in the South Asia region and suggested the SAARC Writers' Forum have its own flag and logo.

Dr. AVS Ramesh Chandra, the political counsellor of the Indian High Commission in Islamabad, on behalf of Indian High Commissioner Shivshankar Menon and his wife Mohini Menon, called for the promotion of goodwill, amity and friendship among the academic and literary community of SAARC member countries.

Mr. Chandra said Indo-Pakistan relations were currently at a moment of hope. "To many of us, it seems we have an opportunity in this transitional moment to move these relations out of the complicated course that they have followed in the past," he said. "It is our unhappy experiences that have lead to some present anxiety and cynicism about the future of the relationship. As a result of the steps towards normalisation that

Which two or more languages can eventually become the test cases?

We are only talking about the "Technical Simple Phrases". The figurative, metaphoric complex literary phrases, are "tight-rope walkers" for machine translation.

If only technical, then do we have the full panoramic view of the technical vocabulary? We estimate it to be nearing one million phrases for each language.

The SAARC writer forum, we have, may be able to develop a platform for our writers for the exchange of information, literary and media related Questions are again so simple:

Do we have correct statistics on these sub-sets; or are we only trying to impose the "IT Lords' Wishful Dreams" onto the common public users?

It needs a new "Esperanto" for the extended "sub-continent". Do we have a sort of Brain-cell" daring enough to tackle this interesting challenge? This situation can avoid English usage as lingua-franca, among the SAARC countries. Urdu may replace this.

The conception phase ended, have we those "genius IT thinkers", who must simultaneously be linguists, to be able to invent that fabulous machine, which can . actualize this intellectual vision... for "invention" it would be!

Paper-book is vanishing and e-book is emerging Are we ready to pace into the speed of this future scenario?

NLA has some good solutions and the' creative minds must move on, to reflect seriously upon these issues. This is the basic thread of the cord of Language and IT twain.

(Read in SAARC Writers Conference, Lahore, March 14, 2004)

enhance the public access to the Govt. information. The Trade activities through Urdu may get a large boost. The software development industry will get into new fields of business activities using Office and Windows in Urdu. This technology will certainly start research activiiies regarding Urdu Informatics, in the universities immediately now and for future studies.

NLA is facing a b~sic challenge, that external character representation of Urdu discourse should be in Nastaleeq and that this Nastaleeq, for memory economy should be letter-based and not image based, as in the case of all existing Nastaleeq softwares which are now in use; even though claims are made that these ligatures are letter-based. That is not an easy job. One has to go for an international support, e.g. UNICODE, Microsoft, ISO etc. With the help of our government, and the vision of NLA succeeded to get full membership of UNICODE. Government of Pakistan, helped NLA to subscribe the membership.

Until and unless there is a separate space for Nastaleeq on UNICODE" software developers should not mislead us that they propose Nastaleeq letter-based Software.

Regarding our future dreams, NLA is now very near to its objectives, i.e. Urdu Spell Check, Urdu Grammar Check, Phraseology, Lexicon, Machine Translation (English to Urdu) etc. The MT facility can be extended by NLA to other Pakistani languages; and also if needed, to the languages of the SAARC countries. Our question is as simple as that:

Do. we have the convenient computational (doubled) grammars to initiate this phase?

If not, how are they to be developed and in how much time?

to select proper words, lexicon, thesaurus etc. Compu poem techniques: to arrive at a rational and proper phrasing of the lines. Springboard and dialogue software:. to develop pleasant dramatical effects. Essays of 3/5 paragraphs: to avoid unnecessary length of our text articles. Fluid text creative manipulation for commercial, > artistic and research demands of the market. Spell-Check, grammar checks etc:. to make our reading printable and transportable in all forms. National Language Authority, Pakistan has already tried to establish all these processes for Urdu writers in particular and our other 36 languages in general.

Starting from ASCII code page up to the Keyboard, NLA included all the Pakistani Languages, which are now shown in the Professional Windows XP.

The MS screen of Windows and Office is also being changed into Urdu, with the collaboration of NLA. This action may lead to create a change of writing habits by an interaction of mind and Keyboard, rather than the older simpler way of thinking. National Language Authority Pakistan is grateful to Microsoft for its big initiative to collaborate in the development of Language for its use in informatics through its "Centum" program Changing the Computer Screen into Urdu will boost up the IT development activities in Pakistan as well I as in India and other SAARC countries and also in Saudi Arabia for Hajj Purposes. It will bring computer in far-flung areas of Pakistan and South Asia and even in every street and village. Teaching of Urdu will be facilitated and education in the schools will be enhanced with the help of computer, because 98% population of Pakistan do not know English. The 2% of English Knowing population has now become a mere small portion of the Literate Persons. Even all the medias will boost their activities e.g. E-mail, Communications and mass media campaigns. The e-Govt. process will be enlarged and will

sequential rhetoric often used in this field, by first looking. into the "future" and then, by analogy, deriving from the domain of the "present", its actual "representation" and "analyses", the "problems" encountered which must be resolved, to attain this predetermined "future".

It is not the hypothetical "should be" that we are looking towards, but the imposing "must" that will enable us to learn the enormous burden that "is" forced on us by the fast onslaught of the MODERN.

As linguists or writers of SAARC languages, we have not considered the discoveries of the many foreign languages. Rather than an entry through the Keyboard by fingers and clicks, we still bank on pen and paper. If once we get acquainted with the Keyboard alphabets for our calligraphy, we may change our thinking style for a better future of Document Engineering.

If Microsoft Office XP was not created, we could not have reached to the point of just a "click" on regional languages. It has been done; still 99% of the computer literates do not use this great contribution i.e. for e-mail, website, word-processing, database and data-banks etc. XP has chosen our (NLA' s) Keyboard, using Arabic script for Urdu browsing/word-processing, to fulfill our computing needs.

Though most of our languages/scripts have space on international code i.e. UNICODE, we have not yet incited the software industry for good resolution of our own needs. Have we not thought constructively on this? It is our obligation to consider the capacity of joining the modem technology and so to encourage and train these inexpert backward minds, for no fault of theirs:

Word processing: to write on Keyboard. Word weavers:

2

Linguistics, IT and Literature

Computer:' the gigantic giant of the Century' has become a nightmare for most of our writers of theTthird World. The horror of English Command in IT related technologies has also made extra pressure on eastern countries. That is why we are still unable to find any commonality in our lingual needs in computer solutions.

Search ideas are so simple in fact that the similarity of these discoveries have so much bewildered the scientific minds, to the extent that their explanations searching the complex has broken the bounds of common-sense Clarifications have thus become almost incomprehensible and un-understandable!

Philology--- It means the study of languages (Past)

Linguistics--- Also means the study of languages (Present)

Unfortunately, by tradition and habit, we tend to rest only on that that we know, so the "Knowledge" in-inertia always hinders us from looking into the "what is of an actual happening; and we deviate into our wishfull thinking and desire thus derogating onto what can be termed "what should be". Most of our supposed derived results are only an expansion of our desires, not a dispassionate statement of scientific analysis.

So to this august meeting, I would like to explain now, what are the problems with the languages of this region in relation to IT, at present. And I will take a different turn to the

the need of the time that an institute of Urdu Informatics in public sector should enhance and boost the localization activities already being rendered in the world of business. The Ex-Cabinet Secretary Mr. Ejaz Raheem (the Health Minister) expressed these view at display of “Pak Nastaleeq” font developed by CEUI on August 20, 2006. He was addressing ceremony before the initial or beta release of Unicode character based highly efficient “Pak Nastaleeq” font. He said that the Prime Minster has decided to move the Urdu quickly into the cyber world. After the successful development of Urdu Office and Urdu Windows in collaboration with the Microsoft, he admired the development of Pak Nastaleeq font which is 90 times faster than other Nastaleeq fonts.

"Institute for Urdu Informatics" will provide a platform to Linguists and IT professionals to initiate more research and efforts in the field of Urdu Informatics.

In a review session in January, 2008 the Cabinet Secretary in chair Mr. Syed Masood Alam Rizvi appreciated the remarkable work done so far in developing Urdu as Computer Language. The future of this work is the R&D and the Institute.

In fact it is voice of this era that an institute for Urdu Informatics should be established so that our graduates may get higher education in the field of Urdu Informatics and it will be provides a strong platform to the research and efforts for National Language to be used in information and communication technologies.

It is a futuristic approach in language development course for Urdu. Urdu Informatics is the future of Urdu and the Institute is the guarantor. We have contributed our share, now what will be the deliberations of our nation?

process on Word/ Office.

5. E- Learning– Prototype Training Courses are to be launched.

6. E- Publishing – Website and Data ware house is being established.

7. Building– Pre qualification/ designing is under way.

In a Cabinet Committee meeting for adoption of Urdu as Official Language on February 1, 2007, it is decided that a project for the Urdu Informatics Institute be prepared. HEC may provide funds for this purpose.

"Urdu Informatics" is a new field or descipline established in IT. It is quoted in Wikipedia, the free encyclopedia that "Urdu Informatics (اُردواطلاعیات) relates to the cutting-edge research and efforts in bringing the utilities and usage of National language to the modern information and communication technologies in education and businesses.

Directorate of Admissions, University of Peshawar has announced that Master, M.Phil and Ph.D in Urdu Course are already in progress in the Department. We intend to start Diploma and Certificate Courses for the people other than from Pakistan, and also working on the Degree Courses of "Urdu informatics" in near future.

In the new Urdu Curriculum for M.Phil & Ph.D; developed by HEC Urdu Informatics or use of Urdu on Computers have been recognized.

Institute for Urdu Informatics is an important and emerging need of the era and it will be soon established. It is

Future Needs of these researches are to be translated as follows:

1. R&D- A Continuous Process for Computational linguistics.

2. Urdu Data House- A Continuous Process- with an Interval of 5 years. (90% Completion is expected in 10 years) for a corpus of 1 billion Urdu Words, including 60 million Urdu verses and 30 million Urdu Phrases.

3. E-Teaching/ Learning- Teaching Courses, Certification, Training.

4. E- Publishing- A Continuous Process.

5. Laboratory Product- Software/ CD's/ Programs, Tools etc.

This type of Institute for Urdu Informatics is needed for R&D and Teaching for this new and emerging field. It started with the project: "Center of Excellence for Urdu Informatics". Its PHASE I was meant for the Development of Standards; PHASE II for the Establishment of Urdu Informatics Center; PHASE III for Urdu Laboratory Products and the Institute.

Its activities are now for:

1. Font Development- Pak Nastaleeq Font is released.

2. Urdu Database– Data Development is being carried out.

3. Machine Translation- for Official Urdu is satisfactorily under construction.

4. OCR- Writing Pad is being developed for Urdu writers to

Both the localization and MT need Computational Linguistics to be researched. Computational linguistics (CL) is a discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences and overlaps with the field of artificial intelligence (AI), a branch of computer science aiming at computational models of human cognition. Computational linguistics has applied and theoretical components. These are localization and translation challenges presented in different Microsoft Conferences e.g. PDC 2006 and 2007. A conference on Language and Technology (CLT 07) was arranged by Peshawar University in August 2007. It revoked that the areas of research are as follows:

- Ambiguity resolution
- Anaphora Resolution
- Character Recognition
- Corpus Linguistics
- Discourse Analysis
- Ellipses Resolution
- Fonts
- Information Retrieval
- Localization
- Machine Translation
- Morphology
- OCR
- Part-of-Speech Tagging
- Pattern Recognition
- Phonology
- Semantics
- Speech Recognition
- Syntax
- Text to speech

(a) Machine Translation Softwares are required for automatic language translation. NLA is in search of a Common Lingual Code to fulfill these requirements.

(b) Steps in M.T. found in English-Urdu lingual code are to go down from complex to simple sentence translation.

1. Lingual Equivalents and Properties/ Diction/ Simple Sentences are to be found or standardized.

2. Clause Equivalents and Rearrangements – Basic Formula for rearrangement of a complex sentence from English to Urdu is devised:

   1/aN+ (n-1) + (n-2) + (n-?) + 1/bV

   Here N=Noun, V= verb, n= total number of clauses. 1, 2, ? are the clauses.

Induction, Gender, Plurality, Conjunctions etc. are also to be taken into considerations and additives are required in editing and polishing of translated Urdu sentences.

Another project of NLA is to develop Urdu Database (Software) for Urdu Databank like English Banks. It may have the following features:

1. Search Items (From idea/ Meaning to word/ phrase).
2. Filing Arrangements.
3. Synonyms and Homonyms.
4. Meaning– Relationship/ Thesauri
5. Data Development
6. Testing
7. Urdu Databank (Corpus) development.

Limerick. There are some other International Localization Institutes also working in this world. PhD Research opportunity with the Localization Research is also available at different universities.

Mobile phone companies like NOKIA, MOTOROLA, SAMSUNG etc. are also localizing their applications. Urdu, Pashto, Sindhi SMS are now in frequent use. A Pakistani company INKSOFT.net made it possible.

NLA's Keyboard is also used by the **google** (Urdu), **Wikipedia.net** (Urdu) Localized.

Localization is now a Global Business. Total Business of Localization in 2005 was $ 8.8 Billion (Urdu is one of the 4 Topmost Microsoft Localization Languages) Microsoft, NOKIA, Motorola, Panasonic, IBM etc. are in this business now. It means that Urdu is already in the mainstream of the IT Industry for localization.

Automatic Language Translation is basic in the field of localization. It is the use of a computer program to translate input text from one national language to another while maintaining the original document format. Yahoo and some other sites offer what is sometimes called instant translation using such a tool. Since language is heavily dependent on context and connoted as well as denoted meaning, a program needs to have access to such context as well as the ability to use it. Since providing enough context is difficult, automatic language translation thus far seems to be successful only in limited and well-understood situations and as a first time-saving step toward translation (or "post-editing") by a human being.

Rhaeto-Romance, Samoan, Tonga,..."

Why Urdu or other Pakistani Languages, be lacking in these positions. What is the remedy? The answer is **Localization.**

- Localization (sometimes shortened to "L10n") is the process of adapting a product or service to a particular language, culture, and desired local "look-and-feel." Ideally, a product or service is developed so that localization is relatively easy to achieve, for example, by creating technical illustrations for manuals in which the text can easily be changed to another language and allowing some expansion room for this purpose.

- In localizing a product, in addition to idiomatic language translation is customized in local culture. A successfully localized service or product is one that appears to have been developed within the local culture.

- Language translation, which is a big and large part of localization, can sometimes be facilitated with automatic language translation.

The Localization Research Centre (LRC) is the information, educational, and research centre for the localization community, established in 1995 at University College Dublin under the Irish Government and European Union funded Technologies Centers programme.

Localization (Education & Research) facilities are now spreading.

New Certified Localization Professional course. MSc. in software Localization Program is also available in University of

the future. Neglecting other languages means ignoring quite significant potential markets.

The most notable feature is the steady rise of Chinese and slow relative decline of Japanese and most European languages. Korean and Indic languages also show growth over that period, though slower than Chinese."

- In general, the data is less reliable for smaller languages, so the order should not be taken as significant.

- Tagalog, Afrikaans, Persian, Swedish, Ukrainian, Malay, Telugu, Greek, Marathi, Tamil, Vietnamese, Cantonese, Urdu, Norwegian, Danish, Czech, Hebrew, Catalan, Romanian, Hungarian, Gujarati, Finnish, Turkic, Punjabi, Kannada, Other Indic, Malayalam, Oriya, Slovak, Galician, Bulgarian, Byelorussian, Croatian, Amharic, Sindhi, Sinhalese, Assamese, Nepali, Kurdish, Kazakh, Uzbek, Slovenian, Pashto (Pushto), Luxembourgish, Azerbaijani, Latvian (Lettish), Cambodian, Turkmen, Basque, Estonian, Albanian, Balochi, Malagasy, Lithuanian, Armenian, Kinyarwanda, Swahili, Laothian, Macedonian, Icelandic, Luri, Georgian, Serbian, Tajik, Hindko, Moldavian, Konkani, Sesotho, Mongolian, Manipuri, Kirghiz, Maltese, Brahui, Chichewa, Croatian, Kirundi, Afar,

ASCII code plate version 2 of NLA. The Ghost Characters set was also recognized and adopted by UNICODE.

There are some more proposals sent by CEUI to UNICODE to develop standards for Urdu and Pakistani languages. The basic concepts of the proposals were as that:

A set of Ghost Characters and dots/ nuqtas makes easy to formulate any character of any language written in Arabic script.

Finally the technical committee of UNICODE has accepted this proposal and in a meeting on October 13, 2007, they were of the opinion that alternate nuqta proposal to add spacing characters be developed. Korean and other CJK languages following this proposal made their contributions. The last leg of this Ghost Theory Proposal was taken as approved by the UNICODE Technical Committee Conveyed on 3 March 2008 and approved in the meeting on 11-12 May 2008. ISO-Open Office Standards also brought for Urdu Standards in its ECMA-376 for Office Open XML file formats. Two meetings of its sub-committees have been conducted.

Why this theory is accepted. This has some basic considerations. A basic rationale is revealed through a research paper of UNICODE developed by Mr. Mark Davis as in GDP by language on January 22, 2003. He writes:

> "Many people in the software industry don't realize how important it is to localize products for different languages around the world. While English is a major language, it only accounts for around 30% of the world Gross Domestic Product (GDP), and is likely to account for less in

companies are also adopting this keyboard.

> "The National Language Authority of Pakistan is also grateful to Microsoft for its great initiative to collaborate in the development of the Urdu Language for use in Informatics through its Local Language Program. The National Language Authority, in collaboration with Microsoft, is working to bring Computer Technology to Urdu. Providing the interface in Urdu will boost IT development activities in Pakistan, as well as in India and other SAARC countries. This initiative will potentially enable computer access in every street and village of Pakistan and South Asia. Teaching of Urdu will be facilitated and education in schools can now be enhanced with the help of computers, because 92 percent of the population of Pakistan does not speak English. In addition, the software development industry will gain a new field of business activities using Office and Windows in Urdu, and we expect that the LLP will also benefit Urdu informatics research activities in Pakistani universities."

This message is also released and available on the website of Microsoft.com as a Press Release of March 16, 2004.

The President of Pakistan also allowed National Language Authority to become full member of International code for language processing named as UNICODE. Some changes are adopted by UNICODE in its Ver. 4.0 from the

Secretary Dr. Masooma Hassan. Under her guidance Ist Urdu Software Computation was held in January 2001.

What were these efforts; let us have a glance of its history. 1st Gazette Notification of Urdu Typewriter's Keyboard was made in 1980. It published in the Gazette of Pakistan, Extra, December 6, 1980 (Pak.III) that: The public sector companies concerned may undertake the manufacturing of Typewriters and Teleprinters with the standardized key boards only.

This keyboard also got a place on Urdu word processing but was not efficient for computing needs. A computing Keyboard Ver.1.00 was developed and released on 14th December 1999 by the Urdu Informatics Department of National Language Authority, Cabinet Division by its standard committee.

This department also developed ASCII Code plate for the use of Urdu in computers. Though the time of ASCII has gone, yet it contributed in UNICODE and ISO standards.

In July 2000 the Chief Executive of Pakistan General Pervaiz Musharaf gave his approval for this standard ASCII code plate. This was revised and the code plate Ver.2 was developed by the standardization committee of NLA. The main feature of this version was Ghost Characters set alongwith dots to process all the Pakistani Languages.

The newly developed keyboard was also made to process according to this new theory. A consensus of Pakistani Language Boards was also taken in November 2000. This keyboard was adopted by NADRA in Pakistan and by Microsoft in Windows XP: English and Urdu versions. Now all other

release their products specifically aimed at Urdu consumer market.

So why on March 2007 the Government of Pakistan in its **Vision 2030** made some lines to make Urdu a language of internet.

1. “True sustainability, however, will come when these languages create their synergies with global modern movements and ideas, especially the Internet.”

2. Urdu is and must remain the first language of Pakistan. It is and must remain the language of our culture and of our day-to-day communication. It is and must remain the first.

The story starts since 1998 when National Identity Cards were proposed to be developed in Urdu but there was a lack of the standardization in Urdu Applications and Softwares. An effort was viewed in a seminar at FAST (NU) Lahore in October 1998 and a resultant Urdu Informatics department started in the National Language Authority, Cabinet Division Islamabad to coop with the running issues of standards.

The Government of Pakistan decided in its cabinet meeting on August 23, 2000 that the development of standards for the use of Urdu for Computer Applications shall continue the responsibility of the Cabinet Division. And after developing so many standards and changing of the computer screen into Urdu, the Prime Minister of Pakistan issued directive, reviewing it on 11-02-2006 as: "The promotion of Urdu language by making it a computer language be put on fast track."

This was all the efforts of support from the ex-Cabinet

society.

2. Hand held Urdu devices will transform our society into a more interconnected society having uncountable benefits.

3. Automatic Machine Translation from English to Urdu will enable the masses to access the data that, before translation, is foreign to them.

4. When it comes to Scientific Text, Automatic Machine Translation will help to translate science and technology literature from abroad into Urdu thus raising the overall intellect of the society.

5. Electronic Urdu Database will help to standardize the language among masses thus boosting the research and refinement of the language which definitely has a positive impact on society.

Advantages to the Economy after completion of such type of the projects may be visualized as under:

1. Computer work in Urdu will encourage less literate businessmen to track their business with digital devices

2. As a result of the previously mentioned consequence, enabling Urdu platforms will boost the local Software industry to produce Urdu Business Solutions.

3. Small business will start to track their business with Urdu enabled devices which in turn will generate both, the business for software industry and revenue for the central government.

4. Urdu platforms will force the Hi-Tech companies to

How this future of Urdu be saved in the era of Computer technology, internet and IT, the answer lies in the development of the projects like Center of Excellence for Urdu Informatics (CEUI) with a view to providing solutions to short-term and long-term language policy, ways and means of adoption of Urdu as Official, Judicial and Instructional language of Pakistan, and conduct research and development for Urdu standardization and academic support to national and international stakeholders and Government of Pakistan.

Advantages to the State with the development of this project are being visualized as:

1. Having everything in Urdu on Computers will promote national integrity more than any other cultural tool.

2. E-Government can take advantage of Urdu Informatics at large by having everything in Urdu accessible for masses.

3. Hand-held Urdu devices will help to cover the state operations to less literate areas where English isn't known.

4. Will add up to a strong localized electronic infrastructure equally accessible from all parts of the state.

5. Automatic Machine Translation will assist the state to keep an eye on world opinions in their own language (Urdu) by continuously translating from various sources.

Advantages to the Society may be reflected as:

1. Urdu Computing platforms will attract masses to these platforms hence eliminating the "digital divide" in our

Processing but instead be expanded to create a database for research and development and internet usage, decided to convert and produce its soft wares in one hundred different languages. Urdu is now on the top of these languages emerged on the digital scene.

Urdu Informatics Department of National Language Authority devised so many projects and standardized many tables which are now in use by the computer world: NADRA, Google, Microsoft, Nokia, Motorola, IBM etc. Changing Computer screen from English-to-Urdu is making it usable for an ordinary Pakistani to operate and use computer easily and locally.

Microsoft Urdu Office 2003, Microsoft Urdu Windows XP and Windows XP Starter Edition have been released by the Microsoft.com.

This initiative will potentially enable computer access in every street and village because 92 percent of the population of Pakistan does not speak or use English.

The future projects of NLA such as Urdu Databases/ Data Bank will also be a great support to all others working in computing.

Before the intervention of NLA, the International Standard for Computers (UNICODE) was adopting Urdu characters and standards erroneously from Arabic code plate and other unreliable sources, especially from India revealing Urdu a subset of Arabic. With NLA now as a Full Corporate Consortium Member of UNICODE for last Eight years, have been effectively warding off interference in Urdu language and diverting its attention towards our needs.

1

Institute for Urdu Informatics:
A Futuristic Approach for Language Development

The year 2008 has been announced by UN as the year of languages for development and dialogues among the languages of the world. As far as our national language Urdu is concerned, it has a two way dialogue openings i.e. with International (English, French, Arabic, Chinese, etc.) and (2) local languages like Hindi, Punjabi, Pashto, Sindhi, Balochi, Siraiki, Hindko, Kashmiri, Brahovi, Shina, Khawar, Balti etc.). There is another vista opened for language development i.e. IT and Computer Science. Future of Urdu language development now lies in this field of Informatics.

By the end of 20th century, it was universally agreed by the linguists and computer technologists that only those languages would survive in the 21st century, which could be effectively used as a Computer Language.

The computer giants like Microsoft, realizing the fact that computer usage can no longer be restricted to Word

0

Urdu Informatics

From Wikipedia, the free encyclopedia

Urdu Informatics (Urdu: اردواطلاعیات) relates to the cutting-edge research and efforts in bringing the utilities and usage of Urdu language to the modern information and communication technologies in education and businesses.

National Language Authority has been at the forefront in introducing Urdu Informatics as a tool for wider standardization of the language. Apart from development of Urdu keyboard, one of the key steps in this respect has been the establishment of a Centre of Excellence for Urdu Informatics in Islamabad, Pakistan.Project Director Dr. Attash Durrani coined this term in 1998 to establish this discipline as a branch of Informatics

Works in Urdu Informatics also relate to the use of Urdu for internet applications, input controls, dynamic association of keyboard buttons to Urdu alphabets and use of proper character coding schemes.

(www.wikipedia.org)

Contents

Urdu Informatics

(Vol:1)

Dr. Attash Durrani

Center of Excellence
for
Urdu Informatics

National Language Authority
Islamabad-Pakistan

اُردو اطلاعیات

(جلد اوّل)

ڈاکٹر عطش دُرّانی